EXHIBIT E

Valuing New Goods in a Model with Complementarity: Online Newspapers

By MATTHEW GENTZKOW*

Many important economic questions hinge on the extent to which new goods either crowd out or complement consumption of existing products. Recent methods for studying new goods rule out complementarity by assumption, so their applicability to these questions has been limited. I develop a new model that relaxes this restriction, and use it to study competition between print and online newspapers. Using new micro data from Washington, DC, I estimate the relationship between the print and online papers in demand, the welfare impact of the online paper's introduction, and the expected impact of charging positive online prices. (*JEL* C25, L11, L82)

The effect of new goods on demand for existing products is often uncertain. Convinced that radio broadcasts were crowding out music sales, record companies in the 1920s waged a series of court battles demanding high royalties for songs, leading some networks to stop playing major-label music altogether (Christopher H. Sterling and John M. Kittross 2001, 214; Paul Starr 2004, 339). It soon became apparent, however, that radio airplay dramatically *increased* record sales, and by the 1950s record companies were paying large bribes to get their songs onto disk jockeys' playlists (Sterling and Kittross 2001, 294).[1] More recently, a much-cited *Business Week* article anticipated that computers would create a "paperless office." Instead, the spread of information technology has sharply increased consumption of paper (Abigail J. Selen and Richard H. R. Harper 2002). Debate continues in the economics literature about the relationships between free file-sharing services and recorded music (Alejandro Zentner 2003; David Blackburn 2004; Felix Oberholzer and Koleman Strumpf 2007; Rafael Rob and Joel Waldfogel 2004), file-sharing services and live concerts (Julie Holland Mortimer and Alan Sorensen 2005), public and private broadcast channels (Steven Berry and Waldfogel 1999; Andrea Prat and David Stromberg 2005), and online and offline retailing (Austan Goolsbee 2001; Todd Sinai and Waldfogel 2004).

Measuring the impact of new goods in such settings is important for several reasons. First, it directly affects firm decisions. A record company's decision to start licensing music for sale online, a publisher's decision to sell the film rights to a novel, a discount retailer's decision to open a new line of more upscale stores, and many other choices about entry, product positioning, and pricing depend critically on the demand-side relationships between new and old products. Estimating these relationships is thus

* Graduate School of Business, University of Chicago, 5807 South Woodlawn Avenue, Chicago, IL 60637 (e-mail: gentzkow@chicagogsb.edu). I would like to offer special thanks to Bob Cohen and Jim Collins of Scarborough Research for giving me access to the data for this study. I thank two anonymous referees for insightful comments. I am also grateful to John Asker, Richard Caves, Gary Chamberlain, Karen Clay, Liran Einav, Gautam Gowrisankaran, Ulrich Kaiser, Larry Katz, Julie Mortimer, Jesse Shapiro, Andrei Shleifer, Minjae Song, and especially to Ariel Pakes for advice and encouragement. Jennifer Paniza provided outstanding research assistance. I thank the Social Science Research Council and the Centel Foundation/Robert P. Reuss Faculty Research Fund at the University of Chicago Graduate School of Business for financial support.

[1] A similar example concerns the introduction of movies. An 1894 article in *Scribners* predicted that the availability of motion-picture and audio versions of novels would lead to the disappearance of printed books (Octave Uzanne 1894). Today, film adaptations and novels are widely perceived to be complements, and film releases are often accompanied by order-of-magnitude increases in sales of the associated book (Kera Bolonik 2001).

Exhibit
PX 2754

important for both firms themselves and economists seeking to understand firm behavior. Second, new goods are a major component of increases in the standard of living, and their omission is a leading source of bias in standard price indices (Timothy F. Bresnahan and Robert J. Gordon 1997). Correcting these biases requires accurate estimates of the effect of new goods on consumer welfare, which cannot be constructed without knowing the relevant demand elasticities. Finally, the degree of substitutability between old and new products is an important input to many policy debates, including those surrounding cable price regulation (Goolsbee and Amil Petrin 2004), deregulation of local phone markets (Robert G. Harris and C. Jeffrey Kraft 1997), and the allowability of cross-media mergers (Federal Communications Commission 2001).

This paper has two goals. First, I extend existing techniques for estimating the impact of new goods to allow for the possibility that goods could be either substitutes or complements. Although a large recent literature studies the effect of new goods,[2] it has been built on discrete-choice demand models whose starting assumption is that consumers choose exactly one product from the set available.[3] This means that all goods are restricted a priori to be perfect substitutes at the individual level. Although this is a reasonable starting point for looking at demand for automobiles or satellite television, it makes these techniques inappropriate for cases such as those described above, where the degree of substitutability or complementarity among products is a key parameter of interest. The new discrete demand model I develop permits consumers to choose multiple goods simultaneously and allows the demand-side relationship between each pair of products to be freely estimated from the data.

Second, I apply the model to study the impact of online newspapers, a good whose relationship with affiliated print newspapers has been hotly debated.[4] I estimate the model using new individual-level data on the print and online newspaper readership of consumers in Washington, DC, and look at the interaction among the *Washington Post,* the *Post*'s online edition (the post.com), and the city's competing daily (the *Washington Times*). I then use the fitted model to ask whether the print and online newspapers are substitutes or complements, and how the introduction of online news has affected the welfare of consumers and newspaper firms. I also address a question of immediate interest to firms: how profits would change if they were to charge positive prices for online content that is currently free.

A central empirical challenge in evaluating the impact of a new good is separating true substitutability or complementarity of goods from correlation in consumer preferences. Observing that frequent online readers are also frequent print readers, that file sharers buy more CDs, or that computer users consume large volumes of paper might be evidence that the products in question are complementary. It might also reflect the fact that unobservable tastes for the goods are

[2] See, for example, Jerry A. Hausman (1997) on the effect of Apple Cinnamon Cheerios, Shane M. Greenstein (1997) on the effect of PCs, Petrin (2002) on the effect of minivans, and Goolsbee and Petrin (2004) on the effect of direct broadcast satellites.

[3] Several existing papers do estimate discrete choice models in which consumers can choose multiple goods. The model developed here differs by allowing goods to range freely from substitutes to complements, and also allowing a flexible form of unobserved consumer heterogeneity. Existing models and their relationship to the present model are discussed in detail below.

[4] According to the *Wall Street Journal,* "Newspaper executives are increasingly debating whether free Web access [to their papers' content] is siphoning off readers from their print operations" (Mike Esterl, "*New York Times* Sets an Online Fee," *Wall Street Journal,* May 17, 2005). See also Leslie Walker, "News Groups Wrestle with Online Fees," *Washington Post,* May 26, 2005; Katharine Q. Seelye, "Can Papers End the Free Ride Online?" *New York Times,* March 14, 2005; and Julia Angwin and Joseph T. Hallinan, "Newspaper Circulation Continues Decline, Forcing Tough Decisions," *Wall Street Journal,* May 2, 2005. Others have argued that an online edition need not crowd out its affiliated print edition and could even complement it (Rob Runnett 2001, 2002). The print-online relationship has been central to the debate surrounding online pricing: "A big part of the motivation for newspapers to charge for their online content is not the revenue it will generate, but the revenue it will save, by slowing the erosion of their print subscriptions" (Seelye 2005). The print-online relationship also looms large in the debate about the long-run viability of print newspapers (Dan Okrent 1999; Gates 2000; David Henry, "Is Buffet too Quick to Write off Newspapers?" *USA Today,* May 4, 2000).

correlated—for example, that some consumers just have a greater taste for news or music overall. In the first section below, I analyze this identification problem in the context of a simple two-good model. I show that the key elasticities are unidentified with data on consumer choices and characteristics alone. I then point out two natural sources of additional information that can aid identification. The first is variables that can be excluded a priori from the utility of one or more goods. In many settings, price is the obvious candidate. The identification argument is also valid for nonprice variables, however, and so can be applied where prices do not vary or where the variation is not exogenous. This is the case in the newspaper market I study, where the price of the online paper is zero throughout the sample. In the estimation, I exploit variables, such as whether consumers have Internet access at work or a fast connection at home, which shift the utility of the online edition without affecting the utility of the print edition.[5] The second potential source of identification is panel data. If correlated unobservables such as taste for news are constant for a given consumer over time, observing repeated choices by the same consumer can allow us to separate correlation and complementarity. For example, a consumer who views the content of two papers as complementary would tend to read both of them on some days and neither on other days. A news junkie who views the papers as substitutes, on the other hand, would also read both with high frequency, but would be more likely to read them on alternate days. In the application, I have data on which newspapers consumers read in the last 24 hours, and also in the last five weekdays, a limited form of panel data I exploit in the estimation.

A further challenge is how to translate the utility estimates from the demand model into dollars. Intuitively, data on consumer choices (combined with exclusion restrictions and panel data) allow us to estimate how consuming one good affects the marginal utility of consuming another. To make welfare statements, we also need to know how consumers trade off these utils of news consumption against dollars. This would be straightforward to estimate if we could observe how demand responds to exogenous variation in prices. I propose an alternative strategy that exploits information from the supply-side of the market and is valid in the absence of price variation. It is based on a simple observation: the less sensitive consumers are to prices, the higher the price a profit-maximizing firm would set for its products. Given observable data on marginal costs and advertising revenue, and a model of the firm's objective function, I can therefore calculate the value of the price elasticity that would equate the profit-maximizing price of the print newspaper with the price we actually observe.[6] This strategy depends on strong assumptions about the form of the firm's profit function, as well as the accuracy of the observed cost data. But sensitivity analysis confirms that the qualitative conclusions are robust to reasonable alternative assumptions.

The results show that properly accounting for consumer heterogeneity changes the conclusions substantially. Both reduced-form OLS regressions and a structural model without heterogeneity suggest that the print and online editions of the *Post* are strong *complements,* with the addition of the post.com to the market increasing profits from the *Post* print edition by $10.5 million per year. In contrast, when I estimate the full model with both observed and unobserved heterogeneity, I find that the print and online editions are significant substitutes. I estimate that raising the price of the *Post* by $.10 would increase post.com readership by about 2 percent, and that removing the post.com from the market entirely would increase readership of the *Post* by 27,000 readers per day, or 1.5 percent. The estimated $33.2 million of revenue generated by the post.com comes at a cost of about $5.5 million in lost *Post* readership. For consumers, the online edition generated a per-reader surplus of $.30 per day, implying a total welfare gain of $45 million per year.

The model also informs the debate about the sustainability of free online content (see footnote 4). I take two approaches to this question.

[5] Zentner (2003) also uses broadband connections as a shifter of Internet use in studying the impact of file sharing on music sales.

[6] Howard Smith (2004) uses a related technique in studying consumer shopping behavior.

The first is to assume that the Post Company may be setting the price of the online edition suboptimally, and ask whether profits could be increased by charging positive prices.[7] I find that, for the period under study, the optimal price is indeed positive, at $.20 per day, and that the loss from charging the suboptimal price of zero is about $8.8 million per year. The second approach is to suppose that the zero price *is* optimal and ask how large transactions costs would have to be to rationalize it. I show that a zero price would be optimal for any transaction cost greater than or equal to $.13 per day. I also show that because of growth in online advertising demand, the gain to raising online prices was virtually eliminated by 2004. This suggests that the zero price may have been part of a rational forward-looking strategy and is approximately optimal today.

Estimating a structural model of the newspaper market is not, of course, the only possible approach to studying the impact of online newspapers. I show below that valuable information can be gleaned by looking at both time series of aggregate newspaper circulation and reduced-form regressions using micro data.[8] There are two major benefits to estimating the complete model, however.[9] First, because the model is derived from utility maximization, it takes on all of the restrictions implied by consumer theory. This means that the estimated parameters can be used to calculate welfare effects. It also allows us to obtain meaningful answers to counterfactual experiments, such as changing the online price, that are outside the variation observed directly in the data. Second, the model allows multiple forms of identification to be brought to bear and combined efficiently in a single estimate. None of the sources of identification I exploit constitutes an ideal natural experiment. Taken together, however, they provide a substantial improvement on the information available in the raw data, lead to sharply different conclusions than would be obtained from naive analysis, and allow us to make progress in understanding a market where the lack of price variation limits the applicability of standard tools.

The next section analyzes the general problem of identifying substitution patterns in a discrete demand model with multiple choices, and provides a brief discussion of related discrete-choice methods. Section II introduces the data and presents reduced-form results on the relationship between print and online demand. Section III specifies the empirical model and estimation strategy, Section IV presents the results, and Section V concludes.

I. Substitution Patterns and Identification

A. *An Illustrative Model*

In this section, I use a simple example to examine identification of substitution patterns in a discrete-choice setting where consumers can choose multiple goods. Suppose there are two goods, labeled A and B, and that consumers can choose at most one unit of each. We observe the choices of a large population of consumers. For simplicity, I will not write the dependence of the model on observable characteristics, assuming that all the consumers in the data are ex ante identical from the econometrician's point of view. The terms below can easily be rewritten as functions of a vector of observables, and the identification arguments interpreted as identification of parameters conditional on this vector.

We can potentially measure three quantities: P_A (the probability of choosing A but not B); P_B (the probability of choosing B but not A); and P_{AB} (the probability of choosing both). The final probability—choosing neither—is linearly dependent so does not provide any additional information.

The goal is to estimate the various own- and cross-price elasticities. These may in turn be inputs into the analysis of the welfare from new goods, the effect of a merger, or the change in profits from offering a different mix of products.

Denote the prices of discrete goods A and B by p_A and p_B. Income not spent on A or B is

[7] Note that the method for calculating the price elasticity described above is based on the assumption that the price of the *print* edition is set optimally. The alternative assumptions I entertain are then (a) that only the zero online price is suboptimal and (b) that all prices are set optimally. These assumptions are discussed in more detail below.

[8] In particular, linear instrumental variables alone provide strong evidence that the print and online papers are substitutes rather than complements (as the raw correlations would suggest).

[9] See also Nevo (2000) and Peter C. Reiss and Frank A. Wolak (2005) for a general discussion of the advantages of structural demand models.

used to purchase a continuous composite commodity. Utility from q units of this commodity is αq which enters overall utility linearly. Denote the utility of consuming a bundle r by u'_r. A natural quantity to define is the double difference:

$$\Gamma = (u'_{AB} - u'_B) - (u'_A - u'_0).$$

This is the discrete analogue of the cross-partial of utility, and measures the extent to which the added utility of consuming good A increases if good B is consumed as well.

Normalizing utility by u'_0, we can define:

$$(1) \qquad \begin{aligned} u_0 &= 0, \\ u_A &= \delta_A - \alpha p_A + \nu_A, \\ u_B &= \delta_B - \alpha p_B + \nu_B, \\ u_{AB} &= u_A + u_B + \Gamma. \end{aligned}$$

Here, $u_r = u'_r - u'_0$, δ_A and δ_B are mean utilities, and ν_A and ν_B represent unobservable variation in utility. I assume that δ_A, δ_B, and Γ are all constant across consumers. Note that these expressions use the fact that the difference between the utility from the composite commodity when good j is purchased ($\alpha(y - p_j)$) and when neither good is purchased (αy) is just $-\alpha p_j$.

To make the discussion concrete, I assume the unobservables are distributed as

$$\begin{bmatrix} \nu_A \\ \nu_B \end{bmatrix} \sim N\left(0, \begin{bmatrix} 1 & \sigma \\ \sigma & 1 \end{bmatrix}\right).$$

The normalization of one of the variance terms to one is without loss of generality, since we can divide all utilities by a constant and not change any of the choice probabilities. The normalization of the other is purely to simplify exposition.

B. *Substitution Patterns*

Let $F(\mathbf{u})$ be the distribution of $\mathbf{u} = (u_A, u_B, u_{AB})$ implied by the assumptions above. Assuming consumers maximize utility, choice probabilities will be given by:

$$(2) \qquad \begin{aligned} P_A &= \int_{\mathbf{u}} I(u_A \geq 0)I(u_A \geq u_B)I(u_A \geq u_{AB})\, dF(\mathbf{u}), \\ P_B &= \int_{\mathbf{u}} I(u_B \geq 0)I(u_B \geq u_A)I(u_B \geq u_{AB})\, dF(\mathbf{u}), \\ P_{AB} &= \int_{\mathbf{u}} I(u_{AB} \geq 0)I \\ &\qquad (u_{AB} \geq u_A)I(u_{AB} \geq u_B)\, dF(\mathbf{u}). \end{aligned}$$

A central focus of this paper will be estimating the degree of substitutability or complementarity among products. Throughout the analysis, I will use the standard modern definition of complements (substitutes): a negative (positive) compensated cross-price elasticity of demand. Note that the definition is not based directly on properties of the utility function (see Paul A. Samuelson 1974 for an extended discussion). I show in this section, however, that in the simple model with two goods there is an intuitive relationship between complementarity and the sign of the interaction term, Γ.

Denote expected demand per consumer for goods A and B by $Q_A = P_A + P_{AB}$ and $Q_B = P_B + P_{AB}$. Because the quasilinear specification of utility causes income to drop out, there are no wealth effects. The elements of the Slutsky matrix are then just the cross-derivatives of demand, and so by the standard definition:

DEFINITION 1: *Goods A and B are substitutes if* $\partial Q_A / \partial p_B > 0$, *independent if* $\partial Q_A / \partial p_B = 0$, *and complements if* $\partial Q_A / \partial p_B < 0$.

Figure 1 shows demand for the goods as regions of (u_A, u_B) space. The first panel shows the case of $\Gamma = 0$, the second panel shows the case of $\Gamma > 0$, and the third panel shows the case of $\Gamma < 0$. To see how the model determines the cross-price derivatives, observe first that increasing p_B is equivalent to shifting probability mass downward. That is, for any point (a, b) in this space, it increases the probability that $u_B \leq b$ given that $u_A = a$.

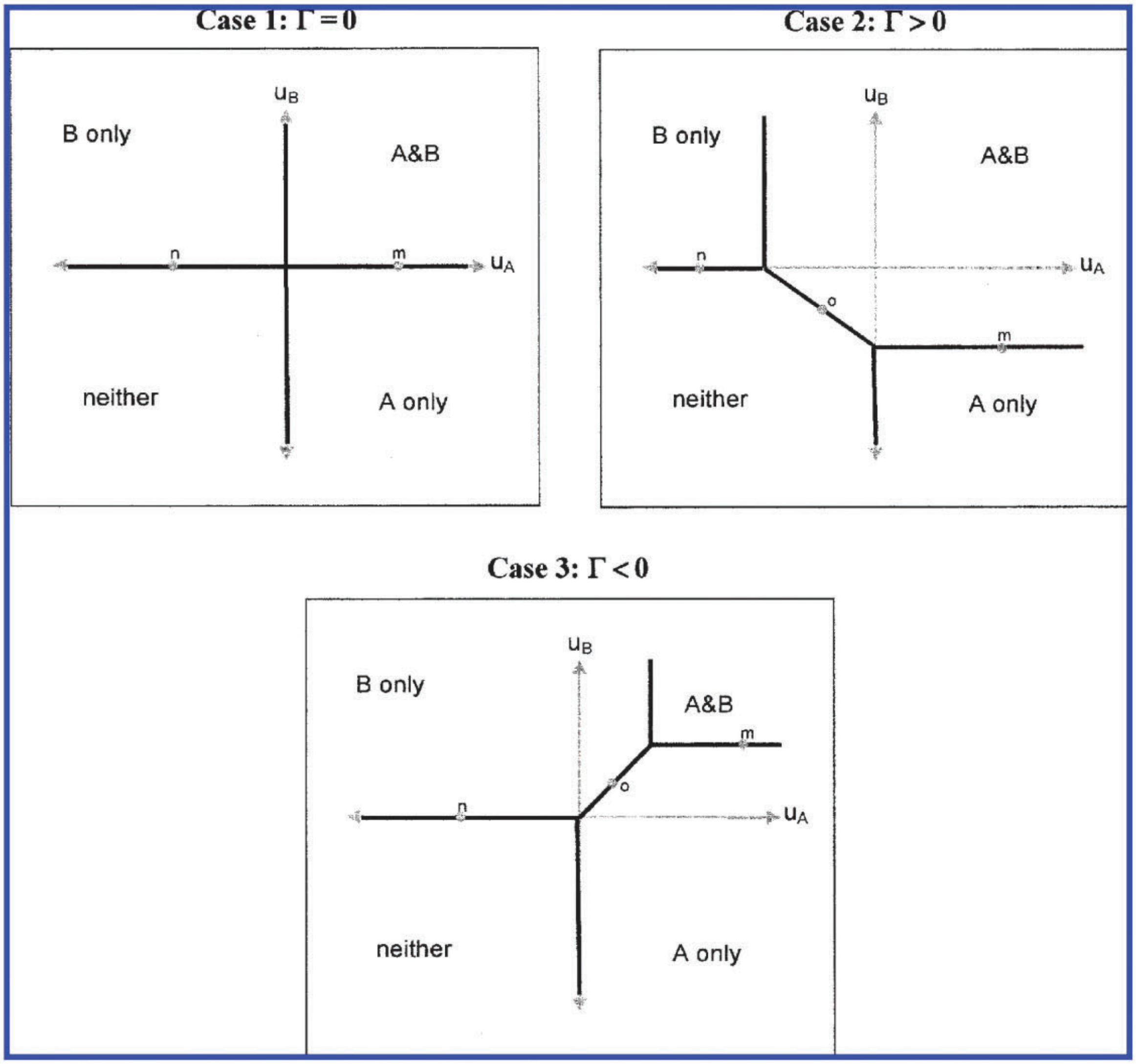

FIGURE 1. ILLUSTRATION OF SUBSTITUTION PATTERNS IN A MODEL WITH TWO GOODS

Notes: Figures show the regions of U_A–U_B space in which the consumer would choose the bundles A and B, B alone, A alone, or neither good. The first panel shows the case where the interaction between the two goods in utility is zero, the second panel the case where it is positive, and the third panel the case where it is negative.

Consider the first panel. Increasing p_B causes marginal consumers such as m to switch from buying the bundle AB to buying A alone. It also causes marginal consumers such as n to switch from buying B alone to buying neither good. Neither of these changes has any effect on the demand for good A, however—the increase in P_A is exactly offset by a decrease in P_{AB}. This implies that when $\Gamma = 0$, the cross-derivatives of demand for the products will be $\partial Q_A/\partial p_B = 0$, and they are therefore independent.

Next, consider the second panel. Increasing p_B causes consumers m and n to switch as before. There will now be consumers such as o, however, who will switch from buying the bundle AB to buying nothing. This means that the drop in P_{AB} will be larger than the increase in P_A, and so $\partial Q_A/\partial p_B < 0$. In the case of $\Gamma > 0$, therefore, the goods are complements.

In the third panel, there are no consumers indifferent between buying AB and buying neither good, but consumers such as o are indifferent between buying A alone and buying B alone.

Increasing p_B causes them to switch from buying B to buying A, so that the increase in P_A is larger than the drop in P_{AB}. We therefore find that $\Gamma < 0$ implies the goods must be substitutes.

This discussion suggests the quite intuitive result that the interaction term Γ is the key parameter for determining the substitutability of goods in a multivariate discrete choice model. Formally, we can substitute into the definition of Q_A and take the derivative with respect to p_B to show that

$$(3) \quad \frac{\partial Q_A}{\partial p_B} = \int_{\mathbf{u}} [I(u_A = u_B)I(-\Gamma \geq u_A, u_B \geq 0)$$

$$- I(u_A + u_B = -\Gamma)I(u_A \leq 0)I(u_B \leq 0)] \, dF(\mathbf{u}).$$

The first term inside the integral represents points on the dark diagonal line segment in the third panel of Figure 1, along which consumers are indifferent between buying A alone and B alone. The second term represents points on the dark diagonal segment in the second panel, along which consumers are indifferent between the bundle AB and buying neither good.

Inspection of equation (3) immediately implies the following result.

PROPOSITION 1: *Goods A and B are substitutes if* $\Gamma < 0$, *independent if* $\Gamma = 0$, *and complements if* $\Gamma > 0$.

While I motivate this result in terms of the thought experiment of changing prices, the application below will be to a situation in which the price of one product—the online paper—is fixed at zero. This does not cause any problems in terms of Definition 1, since a price change around zero is well defined. Furthermore, because utility is quasi-linear, the sign of the cross-price derivatives will be the same as the cross-derivatives with respect to other components of utility. This means we could run through the same intuition from Figure 1 for a shift in nonprice dimensions of utility. Suppose, for example, that good A is a print paper and good B is an online paper. Increasing the utility of B by improving connection speed or making the Internet available at work will shift probability mass in the figure upward just as reducing p_B would, and the effect of such a change on Q_A will be determined by Γ.

For clarity of exposition, this example was restricted to the case of two goods. Gentzkow (2005) shows how the intuition extends to the multi-good case. The situation becomes more complex in a way analogous to standard (continuous) demand theory, but an intuitive link between interaction terms in utility such as Γ and substitution patterns continues to hold.

C. *The Outside Option*

The interpretation of the outside good in this setting is different from its interpretation in the standard multinomial model. In the standard case, the utility of consuming none of the modeled goods—typically indexed as choice zero—is implicitly maximized over all goods excluded from the model. If we are modeling demand for cars, for example, the utility of good zero for consumer *i* would capture the utility of that consumer's best non-car transportation option. It would be the maximum of utility from taking the bus, riding the subway, walking, and so forth.

In a model where choosing multiple goods simultaneously is possible, on the other hand, *all* choices in the model include such an implicit maximization. In the newspaper application, the data do not include consumers' consumption of many news sources, such as cable television, radio, Yahoo! news, and so forth. When a consumer in the data is observed to have read the *Washington Post* on a particular day, it may be that the *Washington Post* was her only source of news on that day, or it may be that she both read the *Washington Post* and watched half an hour of CNN. What the econometrician observes is that the maximum utility of bundles that include only the *Post* is greater for this consumer than the maximum utility of bundles that include any other combination of the observed goods.

One might ask how these unobserved goods will affect the estimated substitution patterns. Suppose, for example, that having watched CNN dramatically reduces the marginal utility of reading the post.com (so that the two are never consumed together) and dramatically increases the marginal utility of reading the *Post* print edition (so that the two are always consumed together).

Suppose, further, that reading the *Post* has no effect on the marginal utility of reading the post.com. From the discussion above, we know that if the *Post* and the post.com were the only two goods in the market, they would be independent in demand. If CNN is present but unobserved, however, we would never see the *Post* and the post.com consumed together, and so would estimate that they are strong substitutes.

What is important to recognize is that the model's answer in both cases would be correct. In a world without CNN, increasing the price of the *Post* would have no effect on demand for the post.com. In a world with CNN, on the other hand, increasing the price of the *Post* would reduce consumption of both it and CNN, which in turn would increase consumption of the post.com. The fact that the true substitutability of a pair of products will depend on both their direct interaction in utility and their indirect interaction via other goods in the market has long been recognized in classical demand theory (Samuelson 1974; Masao Ogaki 1990). The data on consumption of the *Post* and the post.com will allow us to estimate accurately their relationship in demand, whether or not we have data on consumption of other related goods. These estimates, however, will still be *conditional on the set of alternative goods available in the market.* The estimates provide the correct quantity for evaluating the effect of a price change on firm profits. The estimated response to removing the post.com from the choice set will also be correct. The effects could change, however, if the prices or characteristics of important unobserved goods changed dramatically, and the data will of course allow us to say nothing about the relationship between the observed and the unobserved products. Note that these latter limitations are shared by all discrete-choice demand models.[10]

D. *Identification*

Under the assumptions made so far, the model is not identified. There are three observable data points and five independent parameters: δ_A, δ_B, Γ, α, and σ.

The price coefficient α is identified from choice data alone if and only if there is variation in prices. To see this, note that all predicted probabilities would be the same if we replace the parameters $(\delta_A, \delta_B, \alpha)$ by $(\delta_A + \alpha p_A, \delta_B + \alpha p_B, 0)$. With two observed price vectors, on the other hand, we gain three additional moments—any one of these would be sufficient to identify α given the other parameters of the model.

In situations where there is no usable variation in prices, α must be inferred by introducing an additional moment from some other source. In the application below, this comes from one firm's first-order condition. Although only the sums $\delta_A + \alpha p_A$ and $\delta_B + \alpha p_B$ are identified from demand data, there will be a unique α such that the first-order condition is satisfied at the observed price.

The remaining issue is how to separately identify the interaction term, Γ, and the covariance of the unobservables, σ. Intuitively, the mean utilities δ_A and δ_B will be identified by the marginal probabilities Q_A and Q_B. The remaining moment in the data will be how often the goods are consumed together (whether P_{AB} is high relative to P_A and P_B). A high value of P_{AB} can be explained by either a high value of Γ or a high value of σ, and there is nothing left in the data to separate these.

Furthermore, Proposition 1 shows that this leaves the substitution patterns in the model severely unidentified. Without some additional information, the same data could be fit by assuming that the goods are nearly perfect substitutes ($\Gamma \approx -\infty$ and σ high) or nearly perfect complements ($\Gamma \approx \infty$ and σ low). A model that "solves" the problem by imposing an ad hoc restriction on one of these two parameters will

[10] A more subtle issue is how the correct functional form of equation (1) will change in the presence of unobserved third goods. Suppose, for example, that there are three goods A, B, and C, but that only consumption of A and B is observed. If the underlying utilities u'_A, u'_B, etc., are linear in price, the terms such as $\max\{u'_A, u'_{AC}\}$ that will actually be estimated will be linear as well. Beyond this, however, there is no obvious relationship between the functional form of utility with and without the implicit maximization over consumption of C. Of course, we really have no more prior information about the functional form u'_A than we do about $\max\{u'_A, u'_{AC}\}$. Also, it is equally true in standard discrete choice models that the "true" functional form of utilities changes in complex ways as we vary the set of outside goods. The question in the current setting as always is whether the functional form is sufficiently flexible to capture the important variation in the data.

be unlikely to provide a basis for reliable inference about any quantity in which substitutability of the goods plays an important role.

There are, of course, many ways that more moments could be added to the data in order to identify the model. I will briefly discuss two that seem likely to arise frequently in practice and will play a key role in the application. I assume that the necessary technical conditions are satisfied such that the model is identified if and only if the number of moments is greater than or equal to the number of parameters.

The first possible source of identification is exclusion restrictions. Suppose, in particular, that there is some variable x which is allowed to enter the utility of one good, making the mean utility of good A, say, $\delta_A(x)$, but does not enter either δ_B or Γ. One obvious candidate is the price of good A. In the newspaper application considered in this paper, there is no price variation, but there are consumer specific observables such as having Internet access at work that affect the utility of online but not print newspapers. Having observations at a second value of such an x (call this new vector x') would add three new moments ($P_A(x')$, $P_B(x')$, and $P_{AB}(x')$) but only one new parameter ($\delta_A(x')$). The model would therefore be formally identified.

Furthermore, the intuitive basis of the identification is quite strong. Suppose, for example, that the goods are frequently consumed together (P_{AB} is high relative to P_A and P_B). If this is the result of a high Γ, the goods are complements, and shifting up the utility of good A by moving x should also increase the probability of consuming good B. If Γ is zero and the observed pattern is the result of correlation, the probability of consuming good B should remain unchanged.[11]

The second possible source of identification is panel data. Extending the model slightly to allow for repeated choices over time, assume that the observables (ν_A, ν_B) are made up of two components—a possibly correlated random effect term ($\tilde{\nu}_A$, $\tilde{\nu}_B$), which is constant within consumers over time, and an additional time-varying component (ε_A, ε_B), which is assumed to be i.i.d. across products and time. In the newspaper application, this model would amount to assuming that unobserved correlation in the utilities of different papers is driven by consumer characteristics such as a general taste for news that are constant over the course of a week, and that the additional shocks that lead consumers to read on Monday but not Tuesday are uncorrelated.

Now, if we observe each consumer's choice at two different points in time, we have increased the number of moments from 3 to 15.[12] Under the assumption that ($\tilde{\nu}_A$, $\tilde{\nu}_B$) is constant over time, this is sufficient for formal identification of the model parameters, including the full covariance matrix of the random effects. Intuitively, the argument is just a variant of the usual one for the identification of random effects from panel data. Suppose again that goods A and B are frequently consumed together. If this is the result of correlated random effects, we should see some consumers likely to consume both and some consumers likely to consume neither, but conditional on a consumer's average propensity to consume each good, the day-to-day variation should be uncorrelated across goods. If it is the result of a high Γ, on the other hand, the day-to-day variation should be strongly correlated—a given consumer might consume both on one day and neither on another day but would be unlikely to consume either one alone.

A special case that will be relevant to the application below is one where the data are not a true panel but include observations on both a single day's purchases and a summary of purchases over a longer period of time. For example, suppose that consumers in the two-good model make choices on two consecutive days. Suppose we observe the actual choice made on day 1, but not on day 2. We also observe two dummy variables d_A and d_B, where $d_j = 1$ if product j was chosen at least once over the two days. This clearly contains less information than

[11] Michael P. Keane (1992) presents Monte Carlo evidence on the role of this kind of exclusion restriction in identifying the covariance parameters in a multinomial probit model. Since a multinomial probit model defined over bundles effectively nests the model of equation (1), this evidence is relevant. He shows that including exclusion restrictions greatly improves the accuracy of the model.

[12] With observations at two points in time, the moments would be the probability of each possible combination of choices over the two periods. When there are 4 choices, this gives 16 possible combinations. The number of moments is one less than this because the probabilities must sum to one.

a true panel would—if both A and B are chosen on day 1, we will have $d_A = 1$ and $d_B = 1$ regardless of the choice on day 2, and the data therefore provide no information on the day 2 choice. On the other hand, if neither good was chosen on day 1, d_A and d_B will tell us what was chosen on day 2 exactly.

Although this is a more limited form of information about choices over time, it can still separately identify the covariance matrix of the random effects and thus distinguish true complementarity from correlation. To see the intuition for this, consider, first, observations on consumers who chose neither product on day 1. The data will allow us to observe exactly what these consumers chose on day 2. If the variance of $(\tilde{\nu}_A, \tilde{\nu}_B)$ is small, conditioning on the fact that they chose neither good on the first day does not change their choice probabilities on day 2—we should expect the latter to be exactly the same as the choice probabilities in the sample as a whole for day 1. If the variance of the random effects is large, on the other hand, the fact that these consumers did not purchase on day 1 would predict that they would also be less likely to purchase on day 2. We can therefore think of these consumers as identifying the variance of the random effects. The correlation term will then be identified by consumers who chose either A or B, but not both, on day 1. For a consumer who chose A only on day 1, we will see $d_B = 1$ if and only if B was chosen on day 2. If the random effects are strongly positively correlated, observing a choice of A on day 1 suggests that the consumer will be relatively more likely to choose B on day 2. If they are negatively correlated, such a consumer should be less likely to choose B on day 2.

E. *Relationship to Past Literature*

The model of equation (1) provides a useful starting point for understanding the existing approaches in the literature to estimating discrete choices when multiple goods are chosen simultaneously. To make the discussion concrete, suppose we have micro data on demand for two goods, A and B. Suppose that the frequency with which the goods are consumed together is high relative to the frequency with which either is consumed alone. I will discuss what these data would imply for several existing approaches in the literature.

One approach is the multiple-discrete choice model pioneered by Igal Hendel (1999) and applied by Jean-Pierre Dubé (2004). These models assume that the data are generated by an aggregation over a number of individual choice problems, or "tasks." For example, Hendel (1999) estimates demand for PCs by corporations. In this case, a task might represent a single employee's computing needs. Each agent chooses a single good for each task, which makes the task-level problem analogous to equation (1) with $\Gamma_{AB} = -\infty$.[13] Because the utility from using a given good in one task does not depend on what goods were chosen for other tasks, aggregating over a large number of these tasks is similar to aggregating over a population of heterogeneous consumers in a standard multinomial discrete choice model. The model therefore restricts the goods to be substitutes.[14]

A second approach is the multivariate probit (applied, for example, by Angelique Augereau, Shane Greenstein, and Marc Rysman forthcoming). Here, consumption of each good is assumed to be driven by a separate probit equation, with errors possibly correlated across equations. This is exactly equivalent to equation (1) with $\Gamma_{AB} = 0$, and so restricts all goods to be

[13] Both papers allow consumers to choose multiple units of each good, so the task-level choice is more complicated than a standard multinomial discrete choice problem. But the utility specification implies that consumers will choose at most one type of good for each task.

[14] A different parametric restriction on the Γ interaction terms underlies the model of Tat Y. Chan (2006). He defines goods to be a bundle of characteristics, and assumes that the utility of a bundle is a function of the sum of each characteristic across the different goods. The bundle consisting of a bottle of Diet Coke and a bottle of Diet Pepsi, for example, consists of two units of the characteristic "cola," two units of the characteristic "diet," and one unit each of the characteristics "Coke" and "Pepsi." Because utility is assumed in the main specification to be concave in the total of each characteristic, it is subadditive across goods, meaning $\Gamma_{AB} < 0$. This would again imply that the products must be substitutes. (Chan does find complementarity among some products in a specification with many goods, which appears to result from indirect substitution effects as described above.)

independent in demand (all cross-elasticities are zero).[15]

A third approach is to estimate a logit or nested logit model defined over the set of all possible bundles. Papers that take this approach include Charles F. Manski and Leonard Sherman (1980) and Kenneth E. Train, Daniel L. McFadden, and Moshe Ben-Akiva (1987). Because each bundle's utility is parameterized separately, the Γ_{AB} term could be estimated freely (although both of these papers restrict the interactions as a parametric function of the goods' characteristics). The unobservables, on the other hand, are either assumed to be uncorrelated (in the case of the logit) or have a correlation structure dictated by the nests, which is too restrictive to allow the kind of correlation implied by equation (1) with $\sigma \neq 0$. Given the hypothetical data, we would expect such a model to find $\Gamma_{AB} > 0$, implying that the goods would be complements.

The main difference between the current framework and those that exist in the literature is thus a more flexible specification of the way goods interact in utility and the correlation of unobservable tastes. The functional forms for observable and unobservable utility that have been used in the past impose strong restrictions on substitution patterns: for a given set of observations, one could choose models from the literature that would imply that the goods are strong substitutes, independent, or strong complements. In certain settings, such assumptions will be justified, and making them has the obvious benefit of allowing the researcher to analyze larger choice sets than the one considered here. In other settings, the necessary prior information is not available, and it will be critical to allow a more flexible structure and address directly how substitution patterns are identified by the data.

[15] The discrete-continuous framework of Jaehwan Kim, Greg M. Allenby, and Peter E. Rossi (2002) also assumes the equivalent of $\Gamma_{AB} = 0$ (that the utility of a bundle is simply the sum of the utilities of the underlying goods). The conclusion that the goods must be independent does not hold here, however, because the utility of the outside composite commodity is allowed to be concave rather than linear. This implies that all goods will be substitutes, though with a single curvature parameter governing all the cross-elasticities as well as the elasticity of total expenditure on the inside goods.

II. A First Look at the Data

A. *The Scarborough Survey*

The empirical analysis is based on a survey of 16,179 adults in the Washington, DC, Designated Market Area (DMA), conducted between March 2000 and February 2003 by Scarborough Research. The Washington, DC, DMA includes the District of Columbia itself, as well as neighboring counties in Virginia, West Virginia, Pennsylvania, and Maryland. The data include a range of individual and household characteristics of the respondents, as well as information on various consumption decisions. Most importantly for the current application, these include an enumeration of all local print newspapers read over the last 24 hours and 5 weekdays, as well as readership of the major local online newspapers over the same periods.

Washington, DC, has two major daily newspapers, the *Washington Post* and the *Washington Times.* The former is dominant: average daily readership of the *Post* was 1.8 million in 2000–2003, compared to 256,000 for the *Times.* The two papers also differ in their perceived political stance, with the *Times* generally thought to be more conservative than the *Post.* The main online newspaper is the post.com, which had an average of 406,000 area readers per day.[16]

I will define the goods in the model to be daily editions of the *Post,* the *Times,* and the post.com. The outside alternative will include other print and online newspapers, other news sources such as television and radio, and the choice not to consume news at all. As noted above, *all* choices in the model represent an implicit maximization over these outside goods—the observed choice to read the *Post* only, for example, includes consumers who

[16] Readership figures are based on the Scarborough survey. Note that these readership numbers are larger than circulation figures for the same papers, reflecting the fact that multiple consumers read each copy. The *Times* also has an online edition, the washingtontimes.com, but its readership is very small and there are only 373 readers in my sample. In practice, this turns out to be too few to accurately estimate utility parameters for the washingtontimes.com, and so I omit it from the analysis.

TABLE 1—SUMMARY STATISTICS

	Scarborough survey	Washington, DC, DMA (census)
N	16,179	4,203,621
Median income	$62,500	$60,774
Black	20.6%	23.5%
Hispanic	6.4%	7.9%
Female	57.9%	52.1%
Age distribution:		
18–29	17.5%	21.3%
30–39	22.6%	23.4%
40–49	22.2%	21.7%
50–59	17.9%	15.8%
60+	19.8%	17.7%
Highest schooling:		
<High school	7.7%	14.4%
High school	47.0%	42.1%
College	27.2%	26.3%
Graduate	18.0%	17.2%

Notes: The Scarborough survey is a randomized sample of residents of the Washington, DC, DMA 18 years of age and older. All census figures refer to the population of individuals 18 years of age and older, except percent black and Hispanic, which are proportions of all residents. Median income is the population-weighted mean of the median incomes of counties in the Washington, DC, DMA.

read both the *Post* and the *New York Times,* or read the *Post* and watch TV news.

Table 1 gives summary statistics for the Scarborough data along with corresponding census figures for the Washington, DC, DMA. The survey is approximately a 0.4 percent sample, and is broadly representative, with some overrepresentation of older, more educated, and more wealthy individuals, and some underrepresentation of minorities. The survey includes sampling weights to correct for this overrepresentation. I will use the unweighted data for estimation, and use weights when I simulate aggregate effects.[17]

[17] In addition to including weights, the raw Scarborough data also correct for respondents who filled out an initial questionnaire but not the longer survey by filling in a small number of these consumers' survey responses using the responses of other consumers matched by demographics. These "ascribed" observations are easy to identify because the probability of two respondents with the same sampling weight matching perfectly on all survey responses by random chance is very low. I omit these observations (about 6 percent of the initial sample) in all estimation, but include them in the policy simulations in order to get the correct match to aggregate demographics.

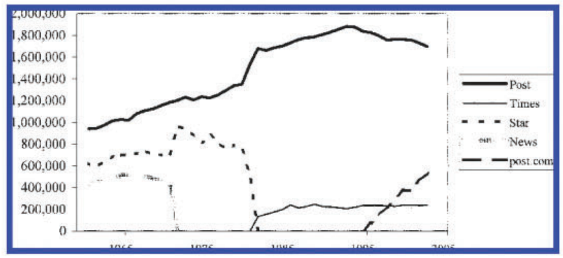

FIGURE 2. READERSHIP OF NEWSPAPERS IN WASHINGTON, DC *(1961–present)*

Notes: Scarborough Research Readership figures are derived by using historical circulation data and the ratio of readership to circulation in the 2000–2003 Scarborough data.
Source: Audit Bureau of Circulations.

B. *Reduced-Form Results*

Figure 2 displays the daily readership of Washington, DC's print and online newspapers since 1961. The first thing to note is that the rapid increase in post.com readership since its introduction in 1996 has been accompanied by a drop in *Post* readership. A simple OLS regression of *Post* readership since 1984 on post.com readership and a time trend gives a significantly negative coefficient, and suggests that it takes four post.com readers to reduce *Post* readership by one. Although it might be tempting to take this as direct evidence that the print and online editions are substitutes, several factors make such a conclusion dubious. First, the downward trend in *Post* readership begins in 1994, two years before the post.com was introduced, and it does not accelerate significantly thereafter. Second, newspaper readership has been declining for many years nationally and there are many demand-side trends that could account for the downward slide of the *Post.* Finally, the downward trend in *Post* readership coincides with a series of increases in the *Post*'s subscription price, and it would be difficult to separate these price effects from the effect of the post.com using aggregate time series alone. For these reasons, getting a handle on the impact of the post.com will require bringing additional information to bear on the problem.

Figure 2 also provides evidence about the extent of substitutability among different print papers. The exit of the *Washington Star* in 1981

TABLE 2—CROSS TABULATION OF *POST* AND POST.COM READERSHIP

24-hour:	Didn't read post.com	Read post.com
Didn't read *Post*	8,771	622
Read *Post*	5,829	877
5-day:	**Didn't read post.com**	**Read post.com**
Didn't read *Post*	6,012	680
Read *Post*	7,203	2,204

and the *Washington News* in 1973 led to increases in the readership of the remaining papers, suggesting some substitutability. In both cases, however, the exit led to declines in total readership, and fewer than half of the readers of the exiting paper appear to have switched to one of the remaining papers. In terms of the *Post* and the *Times,* the time-series provides no evidence of a negative relationship. A linear regression of *Post* readership on *Times* readership actually gives a positive coefficient (though insignificant), even when a time trend is included. Of course, these regressions do not distinguish substitutability from changes in demand or characteristics of the products over time.

Turning to the Scarborough micro data, the first thing to note is that readership of multiple papers is common. Forty-eight percent of consumers reported reading at least one of the *Post, Times,* or post.com in the last 24 hours. Of these consumers, 18 percent reported reading two of the papers, and 1 percent reported reading all three. Over a five-day window, 65 percent of consumers read at least one of the papers; of these, 27 percent read two papers and 3 percent read all three. Table 2 reports the number of consumers reading the *Post* and the post.com over 24-hour and 5-day windows. It is immediately clear from this table that combined readership of print and online news is common. In fact, the fraction of online readers who read print is higher than the fraction of those who do not read online.

Table 3 reports raw and partial correlation coefficients for each pair of papers. The partial correlations control for age, sex, education, industry of employment, employment status, income, political party, date of survey, location of residence within the DMA, and number of missing values in the survey.[18] Readership of the *Post* and post.com are significantly positively correlated over both 24-hour and 5-day windows. Controlling for observable characteristics reduces this correlation by about two-thirds, but it remains significant at the 0.1 percent level. The correlation between readership of the *Post* and the *Times* is also significantly positive in the raw data, but this disappears when controls are added. The partial correlation is zero over a 24-hour window, and significantly negative over a 5-day window. The correlation between the *Times* and the post.com is never significantly different from zero.

What can we conclude from these results? The basic fact in the raw data is that a consumer who reads any one paper is on average more likely to have also read a second paper. If all heterogeneity in utilities were uncorrelated across papers, this would be strong evidence that all three are complements. An alternative explanation is that the kind of consumers who get a lot of value from reading the *Post* also get a lot of value from reading the post.com and the *Times.* The fact that the positive correlation decreases dramatically when we partial out the effect of observables provides direct evidence for this. The question is whether the remaining correlation—in particular the positive correlation between the *Post* and the post.com—represents true complementarity or additional correlation in tastes which is unobserved.

To separate these stories, I will exploit variables that should have a strong effect on the utility of reading the online newspaper, but should have no direct effect on the utility from reading in print. First, I include a dummy variable measuring whether the consumer has Internet access at work. Being able to access the Internet at work clearly reduces the time cost of reading online, but should not directly affect the utility from reading in print. Second, I include two dummy variables indicating whether the consumer uses the Internet for either work-related or education-related tasks. Performing these tasks should

[18] These correlations drop consumers for whom either print or online papers were excluded from the choice set as discussed in the demand specification below.

TABLE 3—CORRELATION COEFFICIENTS

	24-hour		5-day	
	Raw	Partial	Raw	Partial
Post-post.com	0.0989**	0.0364**	0.1579**	0.0673**
Post-*Times*	0.0632**	0.0035	0.0450**	−0.0623**
Times-post.com	0.0146	0.0090	0.0184	0.0066

Notes: The table displays correlation coefficients between dummy variables for reading the *Post,* post.com, and *Times.* In the first two columns, the variable is equal to one if a respondent read in the last 24 hours. In the second two columns, the variable is equal to one if a respondent read in the last five weekdays. Partial correlations are correlations in the residuals from regressions of each consumption dummy on controls for age, sex, education (four categories), white-collar worker, computer worker, employment status, income, political party, date of survey, location of residence within the DMA (six categories), and dummy variables for the number of missing values. Observations where either print or online newspapers were not in the choice set (consumer reports that she generally reads no newspaper sections or did not use the Internet in the last 30 days) were dropped.

** Significant at 1 percent.

lead consumers to be more familiar with the Internet and spend more time at their computers, both of which should decrease the effective cost of reading news online, but not directly affect the utility of print reading. Finally, I include a dummy variable indicating whether the consumer has a high-speed Internet connection at home. This, too, should increase the utility from reading online without directly affecting the utility from reading in print.

Note that an important limitation of the data is that I do not have variables that could be assumed to shift the utility of the *Post* print edition but not the *Times* print edition, or vice versa. I discuss below how the model is identified despite this limitation. In the robustness section, I also show that the results remain qualitatively unchanged when the *Times* is excluded from the analysis completely, suggesting that limited identification on the print side does not bias the estimates of the *Post*-post.com relationship. The reader should bear in mind, however, that the lack of such excluded variables means that the print-print substitution patterns should be interpreted with more caution than the print-online substitution patterns.

One way to see the effect of these excluded variables and perform some checks on their validity is to use them to instrument for online reading in a linear probability model of print reading. There are several problems with such a specification: it does not restrict probabilities to be between zero and one, it restricts the cross-derivative between print and online to be the same for all consumers, and it does not use information from the other choice equations. It shows in an intuitive way, however, how the exclusion restrictions contribute to identification.

The first column of Table 4 shows estimates from a linear probability model of readership over the five-day window, using the same controls as in the partial correlations above. Reflecting the positive correlation noted earlier, the first column shows that reading the post.com is positive and significant in an OLS regression. The second column presents two-stage least squares (2SLS) estimates using the excluded variables as instruments. The coefficient on online reading becomes significantly negative. The magnitude suggests that if we could do an experiment and randomly assign individuals to read the online paper (at zero time cost), they would be on average 40 percent less likely to read the print paper, though the limitations of the linear probability model mean this magnitude must be interpreted with caution. The F statistic on the instruments in the first stage is 33.05, suggesting that weak instruments are not a problem (James H. Stock and Motohiro Yogo 2002). The χ^2 statistic for the overidentification test in this regression is 3.65, with a p value of 0.302, meaning the validity of the instruments cannot be rejected.

A possible concern is that even if the excluded variables do not affect the utility of reading print newspapers directly, they might be correlated with

TABLE 4—LINEAR PROBABILITY MODEL OF *POST* CONSUMPTION

	OLS	IV		
		(1)	(2)	(3)
Dependent variable: Read Post *last 5 days*				
Read post.com last 5 days	0.0464**	−0.4132**	−0.4579**	−0.4381**
	(0.0090)	(0.107)	(0.119)	(0.141)
Other Internet news				0.0244
				(0.0195)
Industry controls			X	
Occupation controls			X	
Overidentification test p value		0.302	0.431	0.219
R-squared	0.333	0.214	0.207	0.202
N	14313	14313	14313	14313

Notes: Robust standard errors are in parentheses. The first row gives coefficients on a dummy for reading the post.com in the last five weekdays. IV regressions instrument for post.com consumption with dummy variables for Internet access at work, fast Internet connection, and reported use of the Internet for research/education and work-related tasks. Overidentification test p value is the p value from a standard Sargan test. Other Internet news is a dummy for online news use other than online newspapers. Industry controls are dummies for 12 industry categories. Occupation controls are dummies for 11 occupation categories. All regressions include controls for *Washington Times* readership, age, sex, education (four categories), white-collar work, computer work, employment status, income, political party, date of survey, location of residence within the DMA (six categories), and dummy variables for the number of missing values. The regressions omit observations where print newspapers are not in the choice set (consumer reports that she generally reads no newspaper sections) and control for presence of online newspapers in the choice set (whether the consumer used the Internet in the last 30 days).

** Significant at 1 percent.

unmeasured components of print utility. Consumers who have a taste for reading news might also have tastes or skills that push them toward the kind of jobs that provide Internet access.[19] An individual who dislikes reading, for example, might be less likely to be successful in white-collar occupations that involve computer-intensive work. Similarly, individuals with a strong taste for news may be more likely to invest in high-speed Internet connections at home. Although it is by definition impossible to test for the presence of such unmeasured correlation, some evidence can be obtained by asking how the 2SLS estimates change when more detailed occupation and industry controls are added. If omitted job characteristics correlated with both Internet availability and taste for reading news were a source of bias, and if these were at least partially captured by occupation and industry controls, we would expect these controls to change the IV estimate. The second IV specification in Table 4 shows how the 2SLS estimate changes when these controls are added. The estimated effect of online news grows slightly stronger, but the basic picture remains the same. This is not proof of the validity of the identifying assumption, but it does provide some confidence that the role of unmeasured correlation is limited (Joseph G. Altonji, Todd E. Elder, and Christopher R. Taber 2005). There is no analogous way to verify the validity of high-speed Internet access as an excluded variable. I show in the robustness section at the end of the paper, however, that the results

[19] A different possibility is that the ability to read Internet news per se is an important determinant of job choice. Because time spent reading online news is a small fraction of the overall time spent at work for most employees, it is unlikely that this is a first-order consideration. One way to verify this is to note that if this channel were important, it would play the biggest role for employees who have recently changed jobs, since Internet access could not have played a role in job choice prior to the mid-1990s. I cannot observe job tenure directly, but I can use as a rough proxy the time respondents have lived in their current area of residence. Limiting the sample to those who have lived in the same area for more than ten years does not significantly change the IV point estimates.

are qualitatively unchanged if this exclusion restriction is not imposed. This suggests that any omitted variables correlated with high-speed access are not significantly biasing the results.

A final concern is that the instruments might drive not only post.com readership but also readership of other online news sources. If these are substitutes for print readership, the IV coefficients could overstate the effect of the post.com. More generally, in order for the exclusion restrictions to be valid, the instruments must not shift the utility of reading print newspapers, either directly *or indirectly* through goods that have strong interactions with the print paper and are included in the outside option. The Scarborough survey asks directly whether consumers use the Internet for news apart from reading online newspapers. As shown in the third IV specification, adding this dummy variable as a control actually increases the magnitude of the post.com coefficient slightly relative to the baseline IV specification in column 2, suggesting that the effect does indeed work through the post.com. Of course, it may be that the instruments affect the intensity of other online news readership in a way not picked up by the dummy variable, in which case we would have to treat the magnitude of the coefficient as an upper bound.[20]

III. Empirical Specification

A. *Demand*

To specify the demand model for estimation, index days by t, consumers by $i = \{1, ..., N\}$, goods by $j \in \{1, ..., J\}$, and the set of possible bundles of these goods by $r \in \{0, 1, ..., 2^J\}$. Assume the bundles are ordered so that $r = 0$ refers to the empty bundle and $r \in [1, J]$ refers to the singleton bundle consisting of only good $j = r$. Define the base utility to consumer i of consuming a single good j on day t to be

$$\bar{u}_{ijt} = -\alpha p_j + \delta_j + \boldsymbol{x}_i \boldsymbol{\beta}_j + \nu_{ij} + \tau_{it}.$$

Here, p_j is the price of good j, δ_j is a constant term, $\boldsymbol{x}_i$ is a vector of observable consumer characteristics, $\boldsymbol{\beta}_j$ is a vector of good-specific coefficients, and ν_{ij} and τ_{it} are unobservables—the former a good-specific shock that is constant across time, and the latter a day-specific shock that affects the utility of all goods equally.

I make several parametric assumptions on the form of the errors. The first error component, ν_{ij}, is meant to capture time-constant consumer characteristics, including overall taste for news (which would show up as positive correlation across the ν_{ij}), brand loyalty to a particular paper, and current subscription status. I assume ν_{ij} has a J-dimensional multivariate normal distribution with a free covariance matrix.[21] The second component, τ_{it}, captures day-to-day variation in the utility of all papers. This could be caused by shocks to the overall quality of news—for example, days when there is a major event such as September 11, 2001—as well as variation in individuals' cost of time. To represent these factors as simply as possible, I assume that τ_{it} has a two-point discrete distribution: it is equal to zero with some probability $(1 - \gamma)$ and equal to a value $\hat{\tau} > 0$ with probability γ.

Because the price coefficient will not be separately identified in the demand estimation, I will replace the first two terms in $\bar{u}_{ijt}$ by a single constant term:

$$\bar{\delta}_j = -\alpha p_j + \delta_j.$$

Later, I will use information from the supply side of the model to identify α and δ_j separately.

To write the utility of a bundle r, define Γ_r to be an interaction term analogous to Γ in equation (1). That is, Γ_r is the difference between the base utility of bundle r and the sum of the $\bar{u}_{ijt}$

[20] In results not reported, I have also verified that the IV coefficient does not change greatly if I add controls for consumption of the *New York Times, Wall Street Journal,* and *USA Today.* As discussed in the outside option section above, the elasticity conditional on consumption of related goods is different from the overall elasticity, and so it would not have been surprising if this coefficient did change. The fact that it does not, however, suggests that indirect substitution effects working through other print papers are small, giving us additional confidence that omission of related products from the complete demand model is unlikely to bias the results.

[21] The variance of the errors will be pinned down by the distribution of the logit errors introduced below, so it is not necessary to impose a normalization on the covariance matrix of the ν_{ij}.

for all goods that comprise bundle r. As a matter of convention, I assume that $\Gamma_r = 0$ for singleton bundles $r \in [1, J]$. Abusing notation to let $j \in r$ denote the set of goods j included in bundle r, I define the utility of consumer i from consuming r on day t to be

$$(4) \qquad u_{irt} = \varepsilon_{irt} \qquad \text{if } r = 0,$$
$$= \sum_{j \in r} \bar{u}_{ijt} + \Gamma_r + \varepsilon_{irt} \qquad \text{if } r > 0,$$

where ε_{irt} is a time-varying shock for each bundle, which will rationalize all remaining day-to-day variation in choices. I assume that the ε_{irt} are i.i.d. type-I extreme value.[22]

Note two important restrictions in this specification. First, I assume that the interaction terms Γ_r do not vary across consumers. I have experimented with specifications that allow Γ_r to vary with observable characteristics, and have found that they do not change the results substantially. Second, because the data lack sufficient price variation to estimate price coefficients that are heterogeneous across consumers, I assume that the price coefficient α is also the same for all i.

The choice set includes three goods: the *Washington Post* print edition, the post.com, and the *Washington Times*. The vector $\mathbf{x}$ includes age, sex, education, industry of employment, employment status, political party, location within the DMA, number of variables on the questionnaire coded as missing, and time dummies coded in six-month intervals which should pick up changes in quality of either the included products or their substitutes.[23] It also includes log income. Although the direct effect of income working through the budget constraint is differenced out in the discrete choice model, there may be important unobservables correlated with income, such as parental education or cognitive skills, which will be proxied by this variable. Dummy variables for Web access at work, use of the Web for education-related tasks, use of the Web for work-related tasks, and broadband connection are allowed to enter utility of the post.com but are excluded from the utility of the two print editions.

Two additional variables define whether either print or online newspapers were part of a consumer's choice set. For print newspapers, I use data from the question, "What newspaper sections do you generally look at?" For the approximately 10 percent of consumers who answered that they did not read any newspaper sections, I assume that print newspapers do not enter their choice set.[24] Similarly, I assume that if a consumer does not report using the Internet at all during the last 30 days, online newspapers are not part of his choice set. I include these variables because I want to condition on any characteristics that affect demand for newspapers and are unlikely to change over the one-week horizon of the data. To allow for the possibility that unobserved characteristics might be correlated with the choice set measures, I include dummies for each as a control in the utility of the goods it does not affect directly (i.e., the "no print" variable enters utility of the online paper, and the "no online" variable enters utility of both print papers).[25]

The model will generate probabilities of each possible choice on a given day. To write this

[22] It is well known that the assumption of extreme value errors is restrictive and can potentially introduce bias in welfare estimates. I show below, however, that the ε_{irt} account for a small portion of the estimated variance in utility. As argued by Petrin (2002), this provides evidence that the extreme value assumption is not biasing the results substantially.

[23] The education dummies are completed high school, completed college, and completed postgraduate degree. Occupation dummies are for white-collar workers and computer workers. Employment status is a dummy for being employed full time. Political party includes dummies for being a registered Republican or Democrat. Dummies for survey periods are each a six-month interval between March 2000 and February 2003. Dummies for location are based on six county-groups comprising the District of Columbia proper, the metro area, and nonmetro surrounding areas. The number of missing observations ranges from 0 to 7 out of approximately 65 questions.

[24] Of consumers who reported reading no newspaper sections, 95 percent also reported reading no print newspapers in the last week. The small number who reported reading no sections but did report reading some paper are dropped from the sample.

[25] Otherwise, substitution patterns would also be identified by variation in the choice set. While this form of identification is valid in many settings, it would be hard to argue here that the choice set variation is exogenous and uncorrelated with other unobserved tastes. It therefore seems safer to include the choice set dummies as controls.

formally, denote the vector of all model parameters by $\boldsymbol{\theta}$. Integrating over the ε_{irt} yields closed-form one-day choice probabilities conditional on the observables and other unobservables:

$$Q_r(\boldsymbol{x}_i, \boldsymbol{\nu}_i, \tau_{it}; \boldsymbol{\theta})$$

$$= \frac{\exp\left[\sum_{j\in r} (\bar{\delta}_j + \boldsymbol{x_i\beta_j} + \nu_{ij} + \tau_{it}) + \Gamma_r\right]}{1 + \sum_{r=1}^{2^J} \exp\left[\sum_{j\in r} (\bar{\delta}_j + \boldsymbol{x_i\beta_j} + \nu_{ij} + \tau_{it}) + \Gamma_r\right]}.$$

I sum over these probabilities to match the form of the data: a dummy variable for whether each product was consumed in the last 24 hours and a separate dummy for whether it was consumed in the last 5 weekdays. Let $\tilde{\mathbf{q}} \in \{0, 1, \ldots, 2^J\}^5$ be a vector indicating the actual bundle chosen on each of the five weekdays covered in the survey. Let q index the possible values of the data we observe—a consumer's reported consumption over the 24-hour and 5-day windows—and let Ω(q) be the (possibly empty) set of $\tilde{\mathbf{q}}$ that are consistent with reported consumption q. The probability of observing consumer i choose q is then

$$P_{\mathrm{q}}(x_i, \nu_i; \boldsymbol{\theta}) = \sum_{\tilde{\mathbf{q}}\in\Omega(\mathrm{q})} \prod_{t=1}^{5} [\gamma Q_{\tilde{q}_t}(\boldsymbol{x}_i, \boldsymbol{\nu}_i, \hat{\tau}; \theta)$$
$$+ (1-\gamma) Q_{\tilde{q}_t}(\boldsymbol{x}_i, \boldsymbol{\nu}_i, 0; \boldsymbol{\theta})].$$

The discussion of the simple two-good model above provides some intuition for the way exclusion restrictions and observations of repeated choices by the same consumer identify the parameters of the demand model. To apply this intuition to the empirical model, note first that the combination of 24-hour and 5-day data will pin down the covariance matrix of the $\boldsymbol{\nu_i}$ random effects. As discussed above, the key point is that the choices in the days before the last 24 hours will be more correlated with the choice in the last 24 hours (conditional on observables) the larger variance of the random effects. For example, knowing an individual read no papers in the last 24 hours would predict she was also less likely to have read in the previous 4 days only if the $\boldsymbol{\nu_i}$ were an important component of utility. Moreover, a positive correlation of the *Post* and post.com random effects would predict that an individual who read only the *Post* in the last 24 hours should be relatively more likely to have read the post.com in the last 5 days. We can then think of the information from the exclusion restrictions as identifying the two parameters of the distribution of the time-varying unobservable, τ_{it}. If the fact that many consumers read both print and online on the same day is driven mainly by τ_{it}, having Internet access at work should not change the probability of reading in print. If this fact is driven by complementarity, those with access at work should be more likely to read.

While this discussion suggests that the model will be formally identified, it is important to emphasize several limitations. First, the identification of the random effects from 24-hour and 5-day data will be weaker than it would be with a true panel. Second, the variables on which we can impose exclusion restrictions shift the utility only of the online paper, which suggests in particular that identification of the substitutability between the two print papers will be weaker than the identification of print-online substitutability. Finally, the data do not include price variation that can be used in the demand estimation. For all these reasons, the imposed functional forms will play a larger role in driving the results than they would in other applications with richer data, and this should be borne in mind in interpreting the findings.

B. *Supply*

As already discussed, limited variation in the prices of the *Post* and the *Times,* and the zero price of the post.com, make it infeasible to estimate the price coefficient α directly. I will, instead, use industry information to approximate advertising revenue and cost parameters, and then use the firm-side pricing equation for the *Washington Post* print edition to back out α.

One complication is how to treat the fact that the *Post* is sold both as single copies and by subscription. I will abstract from the dynamic choice problem this implies for consumers, and assume each pays a single price per copy. (The effect of subscriptions on the choice to read or not will be captured by the $\boldsymbol{\nu_i}$ unobservables.) The cover price of the *Post* increased from \$.25

at the beginning of the sample period (late 2000) to \$.35 at the end of the period (early 2003). The subscription price increased from \$11.16 to \$12.60 for four weeks of daily home delivery. Pricing out Sunday editions at their cover price of \$1.50, the implied subscription price per daily copy ranged from \$.22 to \$.25. I will use the average cover price of \$.30 in the estimation. In the final section, I show that the results are robust to varying the assumed price between \$.25 and \$.35.[26]

Another distinction that has not arisen yet, but will be important here, is the difference between the readership of the paper and the number of copies sold. Comparing the micro data used in this study (which measures readership) to circulation figures from the Audit Bureau of Circulations suggests that the average issue of the *Post* is read by 2.4 adults.[27] This may be accurate, considering the large fraction of papers delivered to multi-occupant households. Alternatively, it may reflect some overreporting in the survey data. It turns out that under plausible assumptions, both scenarios would have the same implications for the model.[28] For simplicity, I will therefore describe the setup of the model assuming there is no overreporting.[29]

Building on these assumptions, I suppose that firm costs are made up of a fixed first-copy cost for both the print and Web editions. I assume that the marginal cost of an additional Web reader is zero in a reasonable neighborhood of the observed number of readers, and the marginal cost of either print edition is constant. Finally, I will use a highly simplified model for advertising demand, namely a constant revenue-per-reader for both print and online. Putting these together, I can specify profits for the Washington Post Company:

$$(5)\quad \Pi = \left[a_p + \frac{p - c_p}{\lambda}\right]N_p + a_w N_w - \Psi,$$

where p is the print edition's price; a_p and a_w are advertising revenue per reported reader for print and online, respectively; N_p and N_w are the number of print and online readers;[30] c_p is the marginal printing and distribution cost per copy; λ is the number of readers per copy; and Ψ is a fixed cost. Based on the assumption that λ represents multiple readership, I assume that the price that enters utility in equation (4) is p_j/λ. In Section IV, I will plug in demand estimates from the model and approximations of the marginal advertising revenue and marginal cost terms taken from industry data. The first-order condition for maximization of Π with respect to the print price will then uniquely define the price coefficient α.

C. *Estimation*

Given the $N \times K$ matrix of $\boldsymbol{x_i}$ ($\mathbf{X}$) and the N-dimensional vector of observed choices of each consumer ($\mathbf{q}$), a natural way to estimate $\boldsymbol{\theta}$ would be to find the value that maximizes the log-likelihood:

[26] Because there is nothing in the data that would allow me to separate these price changes from demand shocks that occurred over the same period, I include time dummies in the model and so do not use this price variation to identify substitution patterns. In the results section below, however, I show that the reactions to these price changes that the model would predict match the actual changes in the time series quite closely.

[27] For comparison, the same calculation for the *Times* suggests that the average issue in 2000 was read by 2.3 adults.

[28] In either case, we would assume that the profit derived from a single reported reader is price minus marginal cost divided by the number of readers per copy, which I denote λ. If the difference between circulation and readership is caused by multiple people reading the same copy, a natural assumption would be that the cost of the issue is shared equally among all readers, meaning the consumer pays p_j/λ rather than p_j. This will mean the estimated α is scaled up by λ, and thus the final utility estimate is scaled down by the same factor. If the data reflect overreporting, we would assume that each reader pays p_j, but we would want to divide the final welfare estimate by λ to take account of the inflated readership numbers. The final answer would thus be the same in either case.

[29] An additional issue related to the fact that consumers do not pay for all the copies that they read could arise if having access to the *Post* without charge were correlated with free access to the *Times*. This would arise if businesses that provided free copies of one also provided free copies of the other. This would be another source of positive correlation between the $\boldsymbol{\nu_i}$ unobservable of the *Post* and of the *Times*. Correlated access to free copies should therefore not bias the results.

[30] Note that N_w may include readers both inside and outside of Washington, DC. I will use the *Post*'s own estimate of the fraction of readership outside of DC (about 25 percent) to scale up the survey-based estimate of DC readership to total readership.

(6)

$$L(\mathbf{X}, \mathbf{q}, \boldsymbol{\theta}) = \sum_i \ln \int_{\nu} P_{q_i}(\boldsymbol{x}_i, \boldsymbol{\nu}; \boldsymbol{\theta})\, dF(\boldsymbol{\nu}; \boldsymbol{\theta}),$$

where $F(\boldsymbol{\nu}; \boldsymbol{\theta})$ is the multivariate normal distribution of $\boldsymbol{\nu}$ conditional on parameters $\boldsymbol{\theta}$. As usual with a random-coefficients model, however, the integral in equation (6) does not have a closed-form solution. I will therefore use simulation draws on the distribution of $\boldsymbol{\nu}$ to form consistent estimates of these probabilities. The simplest way to do this would be to average the conditional probabilities, $P_{q_i}(\boldsymbol{x_i}, \boldsymbol{\nu}; \boldsymbol{\theta})$, over S draws ν_{is} from the multivariate normal distribution of $\boldsymbol{\nu}$ for each consumer i. In the actual estimation, I use an importance-sampling variant of this simulator to generate approximations of the true integral. The simulator reweights the normal distribution to increase the likelihood of drawing ν_{is}, for which the observed choices are relatively likely.[31]

An extensive literature considers techniques for constructing simulation estimators in discrete choice models. The two leading approaches are the simulated maximum likelihood (SML) estimator first discussed by Steven R. Lerman and Manski (1981) and the method of simulated moments (MSM) proposed by McFadden (1989) and Ariel Pakes and David Pollard (1989). The main advantage of the latter is that it is consistent for a fixed number of simulation draws as N goes to infinity, while the former also requires the number of simulation draws, S, to approach infinity with $\sqrt{S}/N = O(1)$. On the other hand, efficiency of MSM requires the optimal instruments to be calculated exactly, and its performance can be quite poor when the preliminary parameter estimate used to generate the instruments is inaccurate (see Christian Gourieroux and Alain Monfort 1996, 43–44, for a discussion). Also, because the MSM estimator requires calculation of the probabilities and derivatives of all possible observed choices, it is computationally costly in cases where the choice set is large—a situation especially likely to arise with panel data. In preliminary tests, I estimated both models and found the parameters did not differ dramatically. Since computation of MSM took 10 to 20 times longer than computation of SML, I will use SML for the final estimation. The SML estimator is

(7)

$$\hat{\boldsymbol{\theta}}^{SML} = \arg\max_{\boldsymbol{\theta}} \left\{ \sum_i \ln \left[\frac{1}{S} \sum_s P_{q_i}(\boldsymbol{x_i}, \nu_{is}; \boldsymbol{\theta}) \right] \right\}.$$

The final estimates are based on 300 simulation draws and adjust equation (7) slightly to incorporate the first-order bias correction suggested by Gourieroux and Monfort (1996, 45).[32] I calculate standard errors using the robust asymptotic approximation proposed by McFadden and Train (2000). Standard errors for welfare estimates and other statistics computed from the model in later stages are estimated by taking 100 draws from

[31] To take importance-sampling draws for a particular consumer i, I use a preliminary estimate of the parameters, $\boldsymbol{\theta}_0$, to generate a simulated approximation of the probability of i's observed choice q_i: $\hat{P}_{q_i}(\boldsymbol{x_i}; \boldsymbol{\theta}_0) = \int_{\boldsymbol{\nu}} P_{q_i}(\boldsymbol{x_i}, \boldsymbol{\nu}; \boldsymbol{\theta}_0)\, dF(\boldsymbol{\nu}; \boldsymbol{\theta}_0)$. Since this is calculated only once for each i, it can be approximated with a large number of simulation draws (I use 200 draws in estimation). I then draw ν_{is} from the distribution

$$(13) \qquad h(\boldsymbol{\nu}|\boldsymbol{x}_i, q_i, \boldsymbol{\theta}_0) = \frac{P_{q_i}(\boldsymbol{x}_i, \boldsymbol{\nu}; \boldsymbol{\theta}_0) f(\boldsymbol{\nu}; \boldsymbol{\theta}_0)}{\hat{P}_{q_i}(\boldsymbol{x}_i; \boldsymbol{\theta}_0)}.$$

The simulated probabilities are formed as a weighted average over S draws from $h(\boldsymbol{\nu}|\boldsymbol{x_i}, q_i, \boldsymbol{\theta}_0)$:

$$(14) \qquad \bar{P}_{q_i}(\boldsymbol{x}_i; \boldsymbol{\theta}) = \frac{1}{S} \sum_s P_{q_i}(\boldsymbol{x}_i, \boldsymbol{\nu}_s; \boldsymbol{\theta}) W_{q_i}(\boldsymbol{x}_i, \boldsymbol{\nu}_s; \boldsymbol{\theta}_0),$$

where $W_{q_i}(\boldsymbol{x_i}, \boldsymbol{\nu}_s; \boldsymbol{\theta}_0) = \hat{P}_{q_i}(\boldsymbol{x_i}; \boldsymbol{\theta}_0)/P_{q_i}(\boldsymbol{x_i}, \boldsymbol{\nu}_s; \boldsymbol{\theta}_0)$. Drawing from $h(\cdot)$ is simple since it is just the distribution of $\boldsymbol{\nu}$ conditional on choice q_i, and draws can thus be taken using an acceptance-rejection method.

[32] The bias correction can be derived as follows. Let $\tilde{f}$ be the simulated value of a choice probability, and let f be the true value. Bias in SML arises because even though $E(\tilde{f}) = f$, it is not true that $E(\ln \tilde{f}) = \ln f$. A second-order expansion of $\ln \tilde{f}$ around f shows

$$E \ln \tilde{f} = \ln f - \frac{1}{2} E \frac{(\tilde{f} - f)^2}{f^2}.$$

This suggests replacing the log term in equation (7) (i.e., $\ln \tilde{f}$) with a consistent estimator of $\ln \tilde{f} + (1/2)E(\tilde{f} - f)^2/f^2$. Writing $P_{is} = P_{q_i}(x_i, \nu_{is})$, the bias-corrected estimator is thus

the asymptotic distribution of the estimated parameters $\hat{\boldsymbol{\theta}}^{SML}$, computing the statistic in question at each draw, and calculating the sample standard deviation (Berry, James Levinsohn, and Pakes 1999; Aviv Nevo 2000).[33]

IV. Results

A. *Demand Parameters*

Table 5 displays SML estimates of the coefficients on observable characteristics. Most of the coefficients in the utility of the *Post* and post.com are significant, as are about half the coefficients in the utility of the *Times*. On the whole, the results correspond closely to expectations.

The coefficients in the utility of the *Post* are consistent with its reputation as a relatively high-brow, liberal newspaper. Both education and income have a positive and significant effect. Considering a consumer with characteristics at the mean of the data, college attendance, graduate school attendance, and doubling household income increase the probability of choosing the *Post* by 31, 34, and 9 percentage points, respectively. Being a registered Democrat increases the probability by 3 points on the margin, while being a registered Republican reduces the probability by 2 points. Age has a positive impact, with an additional 10 years of age adding 8 points to the probability. Being employed full time decreases the probability of choosing the *Post* by 14 points, having a white-collar job not related to computers decreases it by 5 points, and having a computer-related job decreases it by 7 points.

Table 5—Parameter Estimates from Full Model: Observable Characteristics

	Post	post.com	*Times*
Age	0.661**	−0.545**	0.688**
	(0.0407)	(0.0628)	(0.122)
Female	−0.473**	−0.423**	−3.20**
	(0.0921)	(0.148)	(0.344)
High school	1.95**	2.87**	1.29
	(0.214)	(0.633)	(0.856)
College	2.54**	4.03**	1.49
	(0.237)	(0.654)	(0.928)
Grad school	2.76**	4.16**	1.19
	(0.252)	(0.670)	(0.964)
Computer job	−0.567*	1.22**	0.322
	(0.223)	(0.323)	(0.682)
White-collar job	−0.447**	0.431*	−0.591
	(0.118)	(0.195)	(0.423)
Full-time	−1.13**	0.935**	−0.044
	(0.141)	(0.227)	(0.450)
Log income	0.709**	0.217	0.934**
	(0.0729)	(0.118)	(0.260)
Democrat	0.217**	0.326*	−0.027
	(0.099)	(0.163)	(0.346)
Republican	−0.193	−0.0538	2.902**
	(0.119)	(0.193)	(0.408)
Constant	6.97**	−1.54	−8.85**
	(0.852)	(1.18)	(1.91)
N	16179	16179	16179

Notes: Standard errors in parentheses. Details of the model are given in the text. Age is measured in units of ten years. High school, college, and graduate school are mutually exclusive categories. Computer and white-collar job are dummies for reported occupations in these categories and are also mutually exclusive. Full-time is a dummy for full-time employment. Democrat and Republican indicate registered members of the parties. Additional model parameters are shown in Table 6. Not shown in any table are dummies for the number of missing observations, location of the respondent's residence within DC, time dummies (in six-month intervals), and having print and online papers in the choice set.

* Significant at 5 percent.

** Significant at 1 percent.

The coefficients of post.com utility are generally of the same sign as the *Post* coefficients, though their magnitudes in terms of marginal effects are smaller (reflecting the lower predicted probability of choosing the post.com overall—16 percent versus 53 percent for the *Post*). Two notable exceptions are age and employment: adding ten years of age *decreases* the probability of a mean consumer choosing the post.com by 1 percentage point; being employed

$$\hat{\boldsymbol{\theta}}^{SML} = \arg\max_{\theta} \sum_i \left\{ \ln\left[\frac{1}{S}\sum_s P_{is}\right] + \frac{1}{2}\frac{\sum_s \left[P_{is} - \frac{1}{S}\sum_s P_{is}\right]^2}{\left[\frac{1}{S}\sum_s P_{is}\right]^2} \right\}.$$

[33] As discussed by Berry, Levinsohn, and Pakes (1999), this method will generally be superior to the alternative of linearizing the estimates in the parameters and then computing standard errors analytically (the delta method). Another alternative would be to estimate all standard errors in the model by bootstrapping. This would add greatly to the computational cost, since it would require estimating all parameters of the model for each subsample of the original data.

TABLE 6—PARAMETER ESTIMATES FROM FULL MODEL: OTHER

Interaction terms		*Excluded variables (coefficient in utility of post.com)*	
Post-post.com	−1.285**	Internet at work	1.357**
	(0.2307)		(0.180)
Post-Times	0.0809	Fast connection	0.146
	(0.2479)		(0.193)
post.com-*Times*	−1.231**	Use for education-related	0.361
	(0.4832)		(0.212)
Nonlinear parameters		Use for work	0.582**
τ	6.846**		(0.222)
	(0.5027)		
γ	0.0454**		
	(0.0179)		

Notes: Standard errors in parentheses. The model also includes a third-order interaction term for the *Post-Times*-post.com bundle which is not reported in the table. Fast connection indicates consumers with DSL, cable modem, or T1 connections at home. Use variables were responses to the question, "In what ways do you use online services?"
** Significant at 1 percent.

full-time increases the probability by 2 points, having a white-collar job increases it by 1 point, and having a computer-related job increases it by 3 points. The age coefficient is consistent with a widely cited belief in the industry that the online edition has the potential to reach out to younger consumers. The employment coefficients reflect the fact that the online edition is frequently read at work.

The coefficients in the utility of the *Times* are consistent with its reputation for being more conservative than the *Post.* Registered Republicans are significantly more likely to choose the *Times,* with a marginal effect of 3 percentage points, which is large viewed relative to the average predicted probability of choosing the *Times,* which is 9 percent.

Table 6 shows estimates of the other parameters in the model. The first section shows the values of the interaction terms (Γ). Both of the print-online interaction terms are significantly negative. The *Post-Times* term is positive but small and not significant. This implies that the post.com is a substitute for both print papers, consistent with the IV reduced-form regressions (the substitution patterns are discussed in more detail below). The second section shows that the τ and γ parameters are equal to 6.8 and 0.045, respectively. This means that on 4.5 percent of days the utility of all three products increases by about seven units—a natural interpretation is that major news stories, which occur relatively rarely, increase demand for news overall. The third section shows coefficients on the variables that were excluded a priori from utility of the print editions, all of which have the predicted signs, and half of which are significant.

TABLE 7—VARIANCE AND COVARIANCE OF CONSUMER CHARACTERISTICS

Covariance of observable utility	*Post*	post.com	*Times*
Post	14.0		
post.com	7.95	8.43	
Times	7.66	4.85	10.2
Covariance of ν unobservables	*Post*	post.com	*Times*
Post	11.0		
post.com	4.14	23.5	
Times	−1.17	5.19	86.1
Variance of τ unobservables			2.03
Variance of ε unobservables			1.64

Note: The table shows the covariance of different components of utility at the estimated parameter values.

Table 7 shows the sample covariance of the estimated utility from observables, the estimated covariance matrix of the unobservables $\boldsymbol{\nu}$, and the variance of the $\boldsymbol{\tau}$ and $\boldsymbol{\varepsilon}$ unobservables. The correlation of utilities for the *Post* and post.com, and for the post.com and the *Times,* are more positive than the observables alone would predict. The variance of the unobservable component of *Times* utility is quite large, reflecting the fact that *Times* consumption is more consistent over days than the small prob-

abilities from the model without unobservables would allow. Finally, comparing the relative magnitudes of the different utility components shows: (a) the variance of the $\boldsymbol{\nu}_i$ is substantial relative to the observables; (b) the $\boldsymbol{\tau}_i$ unobservables contribute relatively little to the variance (although the difference in utility on the occasional "major news days" when $\boldsymbol{\tau}_i$ is positive is substantial); and (c) the role of the $\boldsymbol{\varepsilon}_i$ is quite small, meaning the usual concerns about incorporating logit errors in estimating the value of new goods are not likely to be a major issue.

One interesting point to note is that the unobserved correlation between the *Post* and *Times* is negative. One possibility is that the print papers are preferred for particular kinds of content—say political coverage—on which tastes in the population are particularly divided.[34] It is also possible that the lack of excluded variables that enter the *Post* and the *Times* separately means that the covariance of their unobservables is underestimated.

The overall fit of the model can be measured in a number of ways. One is to look at the fit of the aggregate predictions. A simulation of choices from the model matches the aggregate shares closely: the overall MSE is 0.0016. We can also look on a more micro level at how well the model is able to fit consumer heterogeneity. One way to see this is to compare the predicted probabilities of a particular choice for consumers who actually consumed that choice to the predicted probabilities for those who did not. The predicted probability of those who made a particular choice is on average 2.1 times as great as the probability for those who did not make that choice.

B. *Supply Parameters and Price Coefficient*

To calculate the marginal advertising revenue a_p and a_w, I assume that the quantity of advertising space both in print and online is fixed (at least over small variations in readership), and the print and online advertising markets are competitive with a fixed price per reader per day. Note that this abstracts from many important features of the advertising market, including differential values for different types of consumers, the extent to which the same reader on two consecutive days is valued differently from two different readers, and possible market power of the *Post.*

I will estimate the advertising parameters by averaging over observed revenue and readership of the *Post* for 2001 and 2002. For 2001 and 2002, the *Post* had total print advertising revenue of \$574.3 million and \$555.7 million, respectively (Washington Post Company 2002). Apportioning this by the percentage of circulation accounted for by the daily edition gives \$1.5 and \$1.4 million per day. For daily circulation of 771,614 in 2001 and 768,600 in 2002, we have an average value of $a_p = \$1.91$.

Online advertising revenue is not made public, but I employ two sources of information to get a ballpark figure. First, total revenue for the *Post*'s online division was \$30.4 million and \$35.9 million for 2001 and 2002, respectively (Washington Post Company 2002), of which the majority was revenue for the post.com. The average daily online readership of the post.com within the Washington, DC, DMA (estimated from the Scarborough data) is 450,457; *Post* financials state that roughly 75 percent of the post.com's total readership is within the DMA. Dividing 75 percent of the average online division revenue by the DC readership gives a per-reader advertising revenue of $a_w = \$.15$. Second, Competitive Media Research tracks online advertising spending for major Web sites. While they do not track the post.com, they provided me with an estimate of June 2001 advertising revenue for the *New York Times* online edition. If we assume that this month was representative, nytimes.com revenue for 2001 was \$32.4 million, which, combined with nytimes.com readership statistics from Media Metrix, yields $\alpha_w = \$.16$.[35] I will use the \$.16 figure in estimation, and then check the robustness of the results

[34] Suppose that there are two types of content: breaking news and political analysis. Breaking news coverage is available only online but consumers prefer reading political analysis in print. Then, if unmeasured political tastes are strong (as they surely are in Washington), we might expect to see liberal news junkies reading the *Post* and the post.com and conservative news junkies reading the *Times* and the post.com but few consumers reading the *Post* and the *Times.*

[35] Competitive Media Research reports that June 2001 advertising revenue for the nytimes.com was \$2,701,085. Media Metrix reports the number of unique readers of the nytimes.com in July 2001 at 5,034,000. Assuming that the ratio of monthly to daily readership at the nytimes.com

to apportioning a smaller fraction of post.com revenue to online advertising.

The next parameter is the marginal cost of printing and distribution for an additional print copy. According to industry sources, the largest component of marginal cost is newsprint. The *Post*'s average annual newsprint consumption in 2001–2002 was 226,796 metric tons (Editor and Publisher 2001), and the average price was \$541 per metric ton (Bureau of Labor Statistics 2005). Using these figures along with the *Post*'s circulation and average number of pages per issue, I estimate that the newsprint cost was \$.37 per daily copy. Considering that there are additional marginal costs of ink and distribution, I will estimate c_p = \$.40 and check robustness to values ranging from \$.30 to \$.50.

The final parameter is the *Post*'s price. As discussed above, I will abstract from the details of subscription and single-copy pricing and assume that each reader faces the same per-copy price. I use the average cover price of \$.30 in the estimation. Below, I show that the results are robust to varying the assumed price between \$.25 and \$.35.

With these estimates, we can now solve for the price coefficient. Note that the additional profit that the *Post* gets from each print reader is:

$$a_p + \frac{p_{post} - c_p}{\lambda} = \$.75,$$

while the additional profit from each online reader is just $a_w = \$.16$. It will be convenient to write the expected number of *Post* print and online readers as a function of $\bar{\boldsymbol{\delta}}$ (the vector of all $\bar{\delta}_j$) and the other estimated parameters $\hat{\boldsymbol{\theta}}$. Denote these by $N_p(\bar{\boldsymbol{\delta}}; \hat{\boldsymbol{\theta}})$ and $N_w(\bar{\boldsymbol{\delta}}; \hat{\boldsymbol{\theta}})$, respectively. Note that it is straightforward to calculate both these quantities and their derivatives with respect to $\bar{\boldsymbol{\delta}}$ for a particular vector of estimated parameters $\hat{\boldsymbol{\theta}}$.[36] Observing that $\partial\bar{\delta}_j/\partial p_j = -\alpha$, we can write the derivative of the Post Company's profit with respect to the price of the *Post* as

$$\frac{\partial \Pi}{\partial p_{post}} = -\alpha\left[\left(a_p + \frac{p_{post} - c_p}{\lambda}\right)\frac{\partial N_p(\bar{\boldsymbol{\delta}}; \hat{\boldsymbol{\theta}})}{\partial \bar{\delta}_{post}} + a_w \frac{\partial N_w(\bar{\boldsymbol{\delta}}; \hat{\boldsymbol{\theta}})}{\partial \bar{\delta}_{post}}\right] + \frac{1}{\lambda} N_p(\bar{\boldsymbol{\delta}}; \hat{\boldsymbol{\theta}}).$$

Given the observed values and model estimates, everything in this equation is known other than α. So long as the term in brackets is nonzero (which will occur at the estimated values with probability one), there is a unique value of α such that the price derivative is zero at the observed value of p_{post}. That is, there is a unique value of α such that the current value of p_{post} maximizes *Post* profits. Following this procedure yields a point estimate of α equal to 9.72.[37]

C. *Substitution Patterns*

One of the main questions that motivated this paper was whether the introduction of the post.com has had a significant crowding out effect on demand for the *Post*. The reduced-form results provided some evidence: in a simple OLS linear probability model, all pairs of goods appeared to be strong complements; in the IV specification, the sign of the *Post*-post.com interaction switched, suggesting

was the same as at the post.com, this implies a daily nytimes.com readership of 562,883.

[36] The expected number of print readers is

$$N_p(\bar{\boldsymbol{\delta}}; \hat{\boldsymbol{\theta}}) = \sum_i \int_{\nu_i} \left\{ \sum_{r \in print} [\gamma Q_{ir}(\boldsymbol{x}_i, \boldsymbol{\nu}_i, \hat{\boldsymbol{\tau}}; \bar{\boldsymbol{\delta}}, \hat{\boldsymbol{\theta}}) + (1-\gamma) Q_{ir}(\boldsymbol{x}_i, \boldsymbol{\nu}_i, 0; \bar{\boldsymbol{\delta}}, \hat{\boldsymbol{\theta}})] \right\} dF(\boldsymbol{\nu}_i; \hat{\boldsymbol{\theta}}),$$

where $r \in print$ denotes the set of all bundles that include the print paper. The number of online readers is defined analogously. To estimate these quantities, I replace the integral over $\boldsymbol{\nu}_i$ with a simulation estimator identical to the one described in the section on welfare estimates below. To estimate the derivatives, I replace the $Q_{ir}(\,)$ terms with their derivatives, which have a closed form.

[37] Note that the assumption here is that p_{post} is chosen optimally conditional on the characteristics of it and all other products in the market. This includes the current zero price of the post.com. The assumption that this price is chosen optimally seems justified by the fact that the subscription price at least has changed frequently over the survey period. Furthermore, I show below that the results do not change substantially if we assume that the "true" price is \$.25 or \$.35 rather than \$.30, suggesting that the estimates will also be robust to small deviations from optimization.

TABLE 8—IMPACT OF THE ONLINE EDITION ON DEMAND FOR PRINT

Case 1: Full model	
Cross-price derivative	8,358
	(1,436)
Change in print readership	−26,822
	(4,483)
Change in print profits	−$ 5,466,846
	(913,699)
Case 2: Model with observable characteristics only	
Cross-price derivative	−8,421
	(752)
Change in print readership	25,655
	(2,270)
Change in print profits	$ 5,229,009
	(462,771)
Case 3: Model with no heterogeneity	
Cross-price derivative	−16,143
	(702)
Change in print readership	51,897
	(2,254)
Change in print profits	$10,577,720
	(459,464)

Notes: Standard errors in parentheses. The table shows three measures of the online edition's impact. The cross-price derivative is the change in post.com readership when the *Post*'s price is increased by $.10. Change in print readership and print profits are the total changes for the *Post* when the online edition is added to the choice set. The table shows the estimated values in three models. Case 1 is the estimates from the full model. Case 2 is a model with observable consumer characteristics but no unobservables other than the i.i.d. logit errors. Case 3 is a model with no observable or unobservable consumer heterogeneity except the i.i.d. logit errors.

substitutability. We can now see how these conclusions hold up in the full model.

The key findings are summarized in Table 8. The table shows three indicators of the degree of complementarity or substitutability between the print and online editions of the *Post:* (a) the cross-price derivative of demand, calculated as the change in readership of the post.com per $.10 change in the price of the *Post;* (b) the change in *Post* readership when the post.com is added to the choice set; and (c) and the cost of this lost readership in terms of profits from the print edition. The first panel shows the results for the full model, the second panel shows the results for a model with observable consumer characteristics only, and the final panel shows the results when the model is estimated allowing for neither observable nor unobservable heterogeneity.

The table suggests two key conclusions. The first is that the print and online papers are significant substitutes. I estimate that a $.10 (33 percent) increase in the price of the *Post* would lead to an increase in post.com readership of 8,358 (2 percent). Comparing actual *Post* demand with the counterfactual simulation in which the post.com is removed from the choice set, I find that introducing the post.com reduces *Post* readership by about 27,000 readers per day and reduces *Post* profits by approximately $5.5 million per year. These effects are all precisely estimated and significantly different from zero. Their magnitude is moderate, however, and certainly does not suggest anything close to a one-for-one crowding out of print readership. The second conclusion is that properly accounting for observed and unobserved consumer heterogeneity is critical to obtaining accurate results. A model that includes consumer observables but no correlated unobservables (i.e., the $\boldsymbol{\nu}_i$ and τ_{it} are fixed at zero) would lead us to conclude that the products are complements rather than substitutes: the cross-price derivative has the same magnitude but the opposite sign of the true estimate, and introducing the post.com is estimated to have *increased* profits from the print edition by $5 million per year. A model with neither observed nor unobserved heterogeneity suggests even stronger

complementarity, with the post.com increasing print profits by $11 million per year.[38]

The relationship of both the *Post* and the post.com to the *Times* is weaker than their relationship to each other. The post.com and *Times* are significant substitutes, and the estimated elasticity implies that a $.10 increase in the price of the *Times* would increase post.com readership by 2,781. The cross-price derivative of the *Post* and the *Times* is negative, close to zero, and insignificant. This implies that the products are essentially independent.

D. *Consumer Surplus*

From the estimated demand parameters, it is a straightforward exercise to calculate the welfare effects of an individual product. For clarity in what follows, I denote the portion of utility excluding the ε_{irt} by $\hat{u}_{ir}(\boldsymbol{\nu}, \tau)$. Let the space of possible bundles be $\mathcal{R}$ and let $\mathcal{R}^{-j} \subset \mathcal{R}$ be the set of bundles that do not include good j. We can then calculate consumer i's expected gain from adding good j to the choice set as:

$$(8)\quad \Delta_i^j = E_{\boldsymbol{\nu},\tau,\boldsymbol{\varepsilon}}[\max_{r\in\mathcal{R}} \hat{u}_{ir}(\boldsymbol{\nu}, \tau) + \varepsilon_{irt}] - E_{\boldsymbol{\nu},\tau,\boldsymbol{\varepsilon}}[\max_{r\in\mathcal{R}^{-j}} \hat{u}_{ir}(\boldsymbol{\nu}, \tau) + \varepsilon_{irt}].$$

Note that this is the expected value from the perspective of the econometrician, given all observed data. Since $\boldsymbol{\nu}$ is constant for a given consumer over time, the choice observed in the data will contain information about the expected value of $\boldsymbol{\nu}$, and the expectation should therefore integrate over $\boldsymbol{\nu}$ conditional on i's choice. The shocks $\boldsymbol{\varepsilon}$ and τ, on the other hand, are i.i.d. across consumers and periods, and observed choices contain no information about its expected value in the future. Thus, the integral over $\boldsymbol{\varepsilon}$ and τ should be unconditional. A standard result on the expectation of the maximum over extreme value errors allows us to rewrite the expectation in equation (8) as

$$(9)\quad E_{\boldsymbol{\nu},\tau,\boldsymbol{\varepsilon}}[\max_{r\in\mathcal{R}} \hat{u}_{ir}(\boldsymbol{\nu}, \tau) + \varepsilon_{irt}] = E_{\boldsymbol{\nu},\tau}\left[\ln \sum_{r\in\mathcal{R}} e^{\hat{u}_{ir}(\boldsymbol{\nu},\tau)}\right].$$

We can estimate the expectation over $\boldsymbol{\nu}$ and τ by simulation

$$(10)\quad E_{\boldsymbol{\nu},\tau}\left[\ln \sum_{r\in\mathcal{R}} e^{\hat{u}_{ir}(\boldsymbol{\nu},\tau)}\right] \approx \sum_{s=1}^{S}\left[\gamma \ln \sum_{r\in\mathcal{R}} e^{\hat{u}_{ir}(\boldsymbol{\nu}_s,\hat{\tau})} + (1-\gamma)\ln \sum_{r\in\mathcal{R}} e^{\hat{u}_{ir}(\boldsymbol{\nu}_s,0)}\right] \times R(\boldsymbol{\nu}_s|r_i),$$

where r_i is i's observed choice, $\hat{\tau}$ is the fitted value of τ, and each ν_s is an independent draw from the estimated (unconditional) distribution of unobservables. The conditional probability of $\boldsymbol{\nu}_s$ given the observed choice r is

$$(11)\quad R(\boldsymbol{\nu}_s|r) = \frac{Q_{ir}(p, \boldsymbol{\nu}_s)f(\boldsymbol{\nu}_s)}{\bar{Q}_{ir}(p)}.$$

TABLE 9—CONSUMER SURPLUS FROM THE POST.COM

Per consumer as % of print	70.4%
Per consumer in $ per day	$0.30
Total as % of print	16.0%
Total in $ per day	$123,413
Total in $ per year	$45,045,862
Standard error	($6,606,939)

Notes: The table shows the loss in consumer surplus that would result if the post.com were removed from the choice set. "Per consumer in $ per day" is the total daily loss divided by the number of post.com consumers. "Per consumer as % of print" is this value as a percentage of the total loss from removing the *Post* print edition from the choice set divided by the number of *Post* readers.

[38] One way to cross-check the validity of the estimated elasticities is to compare the estimated own-price elasticity to an aggregate time series of the *Post*'s price and readership over a longer period. I have calculated a price series for 1994–2003 using a weighted average of the *Post*'s single-copy and subscription prices. I then calculated counterfactual readership, beginning in 1994 and assuming no change in demand other than reaction to price changes at the estimated own-price elasticity of 0.37. The fit of the counterfactual to actual circulation is good, suggesting that *Post* demand is indeed relatively inelastic, and the technique of backing out the price coefficient from the firm's first-order condition yields reasonable results.

TABLE 10—CHANGES IN READERSHIP WHEN POST.COM IS REMOVED

Change in *Post* readership	26,822
Standard error	(4,483)
Consumers who read post.com and *Post*	
Switch to outside good	17.2%
Switch to *Post* only	79.2%
Switch to other bundles	3.5%
Consumers who read post.com only	
Switch to outside good	69.1%
Switch to *Post* only	27.1%
Switch to other bundles	3.7%

Notes: The table shows results from a counterfactual simulation where the post.com is removed from the choice set. The final rows show the distribution of new choices in the counterfactual by consumers whose observed choices were either the *Post*-post.com bundle or the post.com alone.

Here, $\bar{Q}_{ir}(p)$ is a consistent estimator of the unconditional expectation of $Q_{ir}(p, \boldsymbol{\nu})$ over $\boldsymbol{\nu}$ which can easily be generated by simulation. The figures in utils can be converted to dollars by dividing by the price coefficient α and then aggregating to find the total change in consumer surplus

$$(12) \qquad \Delta CS(j) = \sum_i \frac{1}{\alpha} \Delta_i^j.$$

A number of caveats should be emphasized about the validity of these figures. First, we don't actually see prices varying over a large range, so the estimates depend heavily on the assumption of quasi-linearity that allows utils to be converted into dollars at a constant rate. It is variation in consumer characteristics that allows us to map out the demand curve against the scale given by the normalized ε_{it} distribution. Alternative error distributions, or specifications that allowed the variance of ε_{it} to differ across bundles, could yield different results. Second, the analysis here ignores the welfare of consumers who may read the post.com outside of the Washington, DC, area. Finally, the model does not take account of changes in the prices or characteristics of other products in response the post.com's introduction. If the added competition from the post.com caused the quality of competing papers to be improved, for example, we would understate the consumer welfare gains.

The results of the consumer surplus calculation are presented in Table 9. We can get some idea of the welfare effects of the post.com independent of the estimated price coefficient by looking at the surplus of post.com readers relative to the surplus of print readers. I find that the average post.com reader's surplus is 70 percent of the average print reader's surplus. Aggregating over readers, this means that the total consumer welfare gain from the post.com is 16 percent of the total print surplus. In dollars, I find that the average post.com reader would be willing to pay \$.30 per day. The total consumer surplus gain from the post.com is \$123,413 per day, or \$45 million per year.

E. *Producer Surplus and Online Pricing*

The calculation of producer surplus is also straightforward. Given estimates for the advertising and marginal cost parameters, the gain to a given company from the addition of a particular product can be calculated by simulating the change in demand when that product is removed from the choice set, and calculating the resulting change in profit from equation (5).

Similar caveats to the above will apply here. Once again, the results will be sensitive to the stylized supply-side model, and especially the values used for advertising revenue. Furthermore, the producer surplus calculations apply only to the welfare of the *Post* and *Times* companies. This neglects a number of other firms that are affected by the post.com. The most notable example is advertisers. Other examples include other newspaper companies, television and radio companies, and other online news providers.

In Table 10, I present a more detailed breakdown of the counterfactual changes in readership when the post.com is removed from the choice set. As already mentioned, *Post* readership would increase by about 27,000 per day. The bottom rows of this table show the change in the choices of two specific groups of consumers: those who read both the post.com and the *Post* and those who read the post.com alone. Of consumers in the first group, almost 80 percent would switch to reading the *Post* alone, while a smaller number would stop reading entirely.[39] Of consumers in the second

[39] To understand why some readers switch to the outside good, recall that the ε_{irt} unobservables make the interaction of goods in utility heterogeneous. This means that, for some

TABLE 11—PRODUCER SURPLUS EFFECTS OF THE POST.COM
(*Before accounting for post.com operating costs*)

Δ *Post* profit	−$5,466,846
Δ *Times* profit	−$2,687,348
Total post.com revenue	$33,150,000
Δ Producer surplus	$24,995,806
Standard error	($1,876,785)

Notes: The first two rows show changes in firm profits caused by adding the post.com to the choice set. The next row shows the estimated advertising revenue generated by the post.com (including consumers both inside and outside of DC). The producer surplus figure is the effect of the post.com on producer profits, before accounting for the operating costs of the post.com.

group, 27 percent would start reading the *Post,* while almost 70 percent would switch to the outside good.

These demand effects are translated into dollars in Table 11. I estimate that the introduction of the post.com costs about $5.5 million a year in lost print profits, compared to the $33.2 million dollars of revenue generated directly. The effect on the *Times* is smaller, with the customers who substitute to the post.com decreasing profits by about $2.7 million. The total gain in producer surplus from the online edition (before taking account of post.com operating costs) is thus $25 million.

Based on these results, can we understand the Post Company's decision to introduce the post.com? The answer, of course, hinges on the operating costs ignored in the previous calculations. While the *Post* does not break out costs separately for the online division, the *New York Times* does. Reported costs of New York Times Digital were $63.5 million and $67.7 million for 2001 and 2002, respectively (New York Times Company 2003). Given the larger scale of the New York Times online operations, we could take the cost figure of $60 million as an upper bound on the costs of the post.com. On the other hand, the *Post* reported in 2001 that its online division was losing money, so a reasonable lower bound would be $30 million. Taking the midpoint of these bounds as a ballpark figure, the net annual effect of the post.com on producer surplus was a loss of roughly $20 million. This makes the decision seem puzzling.

On the other hand, online revenues almost doubled between 2002 and 2004, driven in large part by an improved market for online advertising. If we use the 2004 revenues (changing only the advertising price per reader but keeping all other parameters the same), the picture changes substantially. The total gain to the Post Company from introducing the post.com is now estimated to be about $56.5 million per year, the effect net of operating costs (again assumed to be $45 million) is a gain of $11.5 million per year, and the total change in producer surplus is an increase of $9 million per year. These calculations suggest that the initial losses of the online paper may have been an investment in anticipation of future market growth.

A final question that the structural model allows us to investigate is whether the zero price of the online edition is surprising. As mentioned in the introduction, I take two approaches to this question. The first is to allow that the Post Company may be setting the price of the online edition suboptimally. Recall that the estimate of the price coefficient was based on the assumption that the price of the print edition *was* set optimally, conditional on the online price. No assumption was made about the optimality of the online price itself. We can therefore ask whether holding product characteristics and the print price constant, the Post Company could increase its profits by charging a positive price for the online edition.[40] I vary the online price, recalculating expected demand for all products, and using this demand to calculate expected profits.

The results show that the estimated profit-maximizing price is $.20 per day, or about $6 per month. Note that because the predicted demand and profits are recomputed from the model for each price, these estimates take account of the direct revenue gains from the price change, the offsetting reduction in advertising revenue, and the resulting increase in print read-

consumers, the products are actually complements, even though they are substitutes in aggregate demand.

[40] Note that because I perform this counterfactual holding the print price constant, the profit-maximizing online price I compute is optimal only in a constrained sense.

TABLE 12—ROBUSTNESS CHECKS

	Post company surplus ($ mil)	Cross-price derivative	Optimal price of post.com	Consumer surplus per year ($ mil)
Actual estimates	$27.68	8,358	$0.20	$45.05
Price = $.25	$27.83	8,598	$0.19	$43.79
Price = $.35	$27.53	8,131	$0.21	$46.31
Marginal cost = $.30	$27.38	7,916	$0.22	$47.56
Marginal cost = $.50	$27.99	8,853	$0.18	$42.53
Online revenue = $25m	$19.53	8,342	$0.24	$45.13
Omit broadband variable	$27.19	9,277	$0.21	$45.12
Only *Post* and post.com	$31.52	8,683	$0.22	$47.15

Notes: The first row repeats the estimates from the actual model that were presented in Tables 5 to 11. The next two rows vary the *Post* cover price used to derive the price coefficient. The following rows vary the value of the marginal cost used. The next row assumes a lower value for the post.com's advertising revenue. The next row omits the fast Internet connection variable that was earlier included in online utility and excluded from print utility. The final row displays estimates from a model with only the *Post* and the post.com (with the *Times* included in the outside good).

ership. The loss from charging a price of zero is approximately $9 million, or more than a quarter of post.com advertising revenue. The $6-per-month estimate is close to the prices of the few online newspapers that did charge for access during the sample period: in 2003, the online subscription prices for two such papers, the *Wall Street Journal* and the *Tulsa World,* were $6.95 and $3.75, respectively.

A second approach is to assume that the Post Company is choosing *both* the print and online prices optimally and use this assumption to estimate a lower bound on the real or perceived transactions cost of online payments. Real costs could arise on the consumer side from the hassle of entering a credit card number or fear of online fraud. For firms, they would be primarily the cost of processing credit card transactions. Consumers might have additional perceived costs if an online newspaper charging positive prices deviates from the prevailing norm and is thus seen as unfair. I assume that the constant term in the utility of the post.com is

$$\bar{\delta}_{post.com} = \delta_{post.com} - \alpha p_{post.com} - \kappa I(p_{post.com} > 0),$$

where κ is the transactions cost and $I(\,)$ is the indicator function. Since $p_{post.com} = 0$ in the observed data, the demand model provides us with an estimate of $\delta_{post.com}$. We can then plug in the earlier computed value of α and look for the lowest value of κ such that total profits are maximized by setting $p_{post.com} = 0$. The results show that κ would have to be at least $.13 per day. Whether this is large or small is a matter of speculation, but it does not seem out of the realm of possibility.

Crucially, the more favorable online advertising market of 2004 greatly reduces the incentives to raise the online edition's price. The optimal price using 2004 online advertising revenue is estimated to be only $.09, and the lost profits from charging a price of zero are just $1.7 million. Furthermore, a zero price would now be optimal for a transaction cost of $.02 or greater.

F. *Robustness Checks*

In Table 12, I present results from the model under a number of alternative assumptions. The first two experiments use values of α derived from the assumption that the price of the *Post* was $.25 and $.35, respectively. Recall that since the former was the price through the end of 2001, and the latter was the price for 2002–2003, the estimates presented earlier were based on the average of these two α values. The table shows that varying α in this range does not substantially change any of the estimates. The next two experiments vary the marginal cost per print copy of the *Post* between $.30 and $.50

($.40 was the value used in estimation). The table shows that the changes in this case are slightly larger—the optimal price varies by $.04, and consumer surplus varies by $5 million—but none of the qualitative results is affected.

The following row uses a lower estimate of post.com revenue than the $33 million assumed earlier. The figure used, $25 million, is a low estimate, taking account of the fact that the *Post*'s online division includes several ventures other than the post.com, including the online edition of *Newsweek*. Predictably, this change does have a significant effect on the estimated surplus from introducing the post.com, reducing it by about a third. Other effects are smaller, however: the estimated optimal price increases by $.04 and estimated consumer surplus is essentially unchanged.

The next row presents results from the model estimated without using data on broadband access. The differences with the preferred estimates are small, suggesting that any correlation between broadband access and unmeasured tastes is not substantially biasing the results.

The final row presents results for the model estimated using data on only *Post* and post.com consumption (thus including the *Times* as an outside good). The results are again qualitatively the same, with the estimated Post Company and consumer surpluses increasing slightly. The fact that the estimated interaction between the *Post* and post.com does not change confirms the theoretical argument made earlier that the estimated substitution patterns will not be biased when important goods are included in the outside alternative. It also confirms that even if the data lack the power to pin down the *Post-Times* interaction perfectly, this will not bias the results of primary interest.

Overall, the results seem robust to varying the supply parameters and form of the utility function, with none of the qualitative results from the model changing under the alternative assumptions considered.

V. Conclusions

Many important questions in economics turn on the extent to which new goods either crowd out or complement existing products. Examples of new goods such as radio, movies, PCs, and file sharing suggest that these relationships can be highly uncertain ex ante. This paper provides new methods for estimating the true relationships and calculating important quantities such as consumer welfare and responses to alternative pricing regimes.

The application to online newspapers addresses three questions. Are print and online newspapers substitutes or complements? How has the introduction of online news affected the welfare of consumers and newspaper firms? And how might demand respond if papers were to charge positive prices for online content that is currently provided free of charge?

I find, first, that print and online papers are clearly substitutes. The apparent positive relationship in the data is an artifact of unmeasured consumer heterogeneity, and disappears in the full demand model. The magnitude of the crowding out of print readership is nonnegligible. It is also small, however, relative to some earlier predictions. Assuming that substitution patterns for newspapers in the Washington, DC, market are broadly representative of substitution patterns elsewhere, the advent of online newspapers does not appear to threaten the survival of print media. Second, the welfare benefits of the online newspaper appear to outweigh its costs. Consumers gain $45 million a year from free provision of the online paper, and although the firm appeared to suffer a net loss during the 2000–2003 period, an improved advertising market means that the current annual effect on firm profits is probably positive. Finally, I find that, in the period under study, the firm could have increased profits by charging a positive price for online content. The potential gain is virtually eliminated at current advertising levels, however, and would disappear with a small transaction cost of online payments.

REFERENCES

Altonji, Joseph G., Todd E. Elder, and Christopher R. Taber. 2005. "Selection on Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools." *Journal of Political Economy*, 113(1): 151–84.

Augereau, Angelique, Shane Greenstein, and Marc Rysman. Forthcoming. "Coordination versus Differentiation in Standards War: 56k Modems." *RAND Journal of Economics*.

Berry, Steven, James Levinsohn, and Ariel Pakes. 1999. "Voluntary Export Restraints on Automobiles: Evaluating a Trade Policy." *American Economic Review*, 89(3): 400–30.

▶**Berry, Steven T., and Joel Waldfogel.** 1999. "Public Radio in the United States: Does It Correct Market Failure or Cannibalize Commercial Stations?" *Journal of Public Economics*, 71(2): 189–211.

Blackburn, David. 2004. "Online Piracy and Recorded Music Sales." Unpublished.

Bolonik, Kera. 2001. "Page to Screen and Back." *Publishers Weekly*. October 1.

Bresnahan, Timothy F., and Robert J. Gordon ed. 1997. *The Economics of New Goods*. National Bureau of Economic Research, *Studies in Income and Wealth*, Vol. 58. Chicago: University of Chicago Press.

Bureau of Labor Statistics. 2005. "Commodity Price Index," Series ID WPU09130291, http://www.bls.gov (accessed March 5, 2005).

Chan, Tat Y. 2006. "Estimating a Continuous Hedonic Choice Model with an Application to Demand to Soft Drinks." *RAND Journal of Economics*, 37(2): 466–82.

▶**Dubé, Jean-Pierre.** 2004. "Multiple Discreteness and Product Differentiation: Demand for Carbonated Soft Drinks." *Marketing Science*, 23(1): 66–81.

Editor and Publisher. 2001. *Editor and Publisher International Yearbook 2001: The Encyclopedia of the Newspaper Industry.* New York: Editor and Publisher.

Federal Communications Commission. 2001. "FCC Initiates Proceeding to Review Newspaper Broadcast Cross-Ownership Rule." Press Release: September 13.

Gates, Dominic. 2000. "E-Book Evangelist: Microsoft's Brash Dick Brass Leads the Revolution that could Bury Dead Trees." *Industry Standard,* September 25. http://www.thestandard.net (accessed November 8, 2002).

Gates, Dominic. 2002. "News Sites Hustle for Profitability." *USC Annenberg Online Journalism Review,* July 11. http://www.ojr.org (accessed November 8, 2002).

Gentzkow, Matthew. 2005. "Valuing New Goods in a Model with Complementarity: Online Newspapers." Unpublished.

Goolsbee, Austan. 2001. "Competition in the Computer Industry: Online versus Retail." *Journal of Industrial Economics*, 49(4): 487–99.

▶**Goolsbee, Austan, and Amil Petrin.** 2004. "The Consumer Gains from Direct Broadcast Satellites and the Competition with Cable TV." *Econometrica*, 72(2): 351–81.

Gourieroux, Christian, and Alain Monfort. 1996. *Simulation-Based Econometric Methods*. CORE Lectures. Oxford: Oxford University Press.

Greenstein, Shane M. 1997. "From Superminis to Supercomputers: Estimating Surplus in the Computing Market." In *The Economics of New Goods*, ed. Timothy F. Bresnahan and Robert J. Gordon, 329–62. National Bureau of Economic Research, *Studies in Income and Wealth*, Vol. 58. Chicago: University of Chicago Press.

Harris, Robert G., and C. Jeffrey Kraft. 1997. "Meddling Through: Regulating Local Telephone Competition in the United States." *Journal of Economic Perspectives*, 11(4): 93–112.

Hausman, Jerry A. 1997. "Valuation of New Goods under Perfect and Imperfect Competition." In *The Economics of New Goods*, ed. Timothy F. Bresnahan and Robert J. Gordon. 209–37. National Bureau of Economic Research, *Studies in Income and Wealth*, Vol. 58. Chicago and London: University of Chicago Press.

▶**Hendel, Igal.** 1999. "Estimating Multiple-Discrete Choice Models: An Application to Computerization Returns." *Review of Economic Studies*, 66(2): 423–46.

▶**Keane, Michael P.** 1992. "A Note on Identification in the Multinomial Probit Model." *Journal of Business and Economic Statistics*, 10(2): 193–200.

▶**Kim, Jaehwan, Greg M. Allenby, and Peter E. Rossi.** 2002. "Modeling Consumer Demand for Variety." *Marketing Science*, 21(3): 229–50.

Lerman, Steven R., and Charles F. Manski. 1981. "On the Use of Simulated Frequencies to Approximate Choice Probabilities." In *Structural Analysis of Discrete Data with Econometric Applications,* eds. Charles F. Manski and Daniel McFadden, 305–19. Cambridge, MA: MIT Press.

Manski, Charles F., and Leonard Sherman. 1980. "An Empirical Analysis of Household Choice among Motor Vehicles." *Transportation Research,* 14(A): 349–66.

▶**McFadden, Daniel.** 1989. "A Method of Simulated Moments for Estimation of Discrete

Response Models without Numerical Integration." *Econometrica*, 57(5): 995–1026.

▶**McFadden, Daniel, and Kenneth Train.** 2000. "Mixed MNL Models for Discrete Response." *Journal of Applied Econometrics*, 15(5): 447–70.

Mortimer, Julie Holland, and Alan Sorensen. 2005. "Supply Responses to Digital Distribution: Recorded Music and Live Performances." Unpublished.

▶**Nevo, Aviv.** 2000. "A Practitioner's Guide to Estimation of Random-Coefficients Logit Models of Demand." *Journal of Economics and Management Strategy*, 9(4): 513–48.

▶**Nevo, Aviv.** 2001. "Measuring Market Power in the Ready-to-Eat Cereal Industry." *Econometrica*, 69(2): 307–42.

New York Times Company. 2003. "The New York Times Company Reports; Fourth Quarter and Full-Year Results." Press release, January 28. http://www.nytyco.com/press.html (accessed October 31, 2003).

▶**Oberholzer, Felix, and Koleman Strumpf.** 2007. "The Effect of File Sharing on Record Sales: An Empirical Analysis." *Journal of Political Economy*, 115(1): 1–42.

Ogaki, Masao. 1990. "The Indirect and Direct Substitution Effects." *American Economic Review*, 80(5): 1271–75.

Okrent, Dan. 1999. "The Death of Print?" *The Digital Journalist*, December 14. http://www.digitaljournalist.org/issue0002/okrent.htm (accessed November 8, 2002).

▶**Pakes, Ariel, and David Pollard.** 1989. "Simulation and the Asymptotics of Optimization Estimators." *Econometrica*, 57(5): 1027–57.

▶**Petrin, Amil.** 2002. "Quantifying the Benefits of New Products: The Case of the Minivan." *Journal of Political Economy*, 110(4): 705–29.

Prat, Andrea and David Stromberg. 1989. "Commercial Television and Voter Information." Unpublished.

Reiss, Peter C., and Frank A. Wolak. 2005. "Structural Econometric Modeling: Rationales and Examples from Industrial Organization." In *Handbook of Econometrics 6*, ed. James J. Heckman and Edward E. Leamer. Amsterdam: Elsevier Science, North-Holland.

Rob, Rafael, and Joel Waldfogel. 2004. "Piracy on the High C's: Music Downloading, Sales Displacement, and Social Welfare in a Sample of College Students." Unpublished.

Runnett, Rob. 2001. "NY Times Digital finds "Magical Mix" in Print/Online Media Relationship." *Newspaper Association of America*. http://www.naa.org (accessed February 26, 2002).

Runnett, Rob. 2002. "Yes, Circulation Directors, They're Still Buying the Paper: Single-Copy Purchases up During Belden Q3 survey." *The Digital Edge,* January. http://www.naa.org/CirculationPages/Yes-Circulation-Directors-Theyre-Still-Buying-the-Paper.aspx (accessed March 10, 2005).

Samuelson, Paul A. 1974. "Complementarity: An Essay on the 40th Anniversary of the Hicks-Allen Revolution in Demand Theory." *Journal of Economic Literature*, 12(4): 1255–89.

Selen, Abigail J., and Richard H. R. Harper. 2002. *The Myth of the Paperless Office.* Cambridge, MA: MIT Press.

▶**Sinai, Todd, and Joel Waldfogel.** 2004. "Geography and the Internet: Is the Internet a Substitute or a Complement for Cities?" *Journal of Urban Economics*, 56(1): 1–24.

▶**Smith, Howard.** 2004. "Supermarket Choice and Supermarket Competition in Market Equilibrium." *Review of Economic Studies*, 71(1): 235–63.

Starr, Paul. 2004. *The Creation of the Media: Political Origins of Modern Communications.* New York: Basic Books.

Sterling, Christopher H., and John M. Kittross. 2001. *Stay Tuned: A History of American Broadcasting*. Mahwah, NJ: Lawrence Erlbaum Associates.

Stock, James H., and Motohiro Yogo. 2002. "Testing for Weak Instruments in Linear IV Regression." Unpublished.

▶**Train, Kenneth E., Daniel L. McFadden, and Moshe Ben-Akiva.** 1987. "The Demand for Local Telephone Service: A Fully Discrete Model of Residential Calling Patterns and Service Choices." *RAND Journal of Economics*, 18(1): 109–23.

Uzanne, Octave. 1894. "The End of Books." *Scribner's Magazine Illustrated*, 16: 221–31.

Washington Post Company. 2002. *Annual Report 2002.* Washington, DC: Washington Post Company.

Zentner, Alejandro. 2003. "Measuring the Effect of Music Downloads on Music Purchases." Unpublished.